

Reeb Graphs: Approximation and Persistence

Tamal K. Dey*

Yusu Wang*

Abstract

Given a continuous function $f : X \rightarrow \mathbb{R}$ on a topological space X , its *level set* $f^{-1}(a)$ changes continuously as the real value a changes. Consequently, the connected components in the level sets appear, disappear, split and merge. The Reeb graph of f summarizes this information into a graph structure. Previous work on Reeb graph mainly focused on its efficient computation. In this paper, we initiate the study of two important aspects of the Reeb graph which can facilitate its broader applications in shape and data analysis.

The first one is the approximation of the Reeb graph of a function on a smooth compact manifold M without boundary. The approximation is computed from a set of points P sampled from M . By leveraging a relation between the Reeb graph and the so-called *vertical homology group*, as well as between cycles in M and in a Rips complex constructed from P , we compute the H_1 -homology of the Reeb graph from P . It takes $O(n \log n)$ expected time, where n is the size of the 2-skeleton of the Rips complex. As a by-product, when M is an orientable 2-manifold, we also obtain an efficient near-linear time (expected) algorithm for computing the rank of $H_1(M)$ from point data. The best known previous algorithm for this problem takes $O(n^3)$ time for point data.

The second aspect concerns the definition and computation of the *persistent Reeb graph homology* for a sequence of Reeb graphs defined on a filtered space. For a piecewise-linear function defined on a filtration of a simplicial complex K , our algorithm computes all persistent H_1 -homology for the Reeb graphs in $O(nn_e^3)$ time, where n is the size of the 2-skeleton and n_e is the number of edges in K .

1 Introduction

Given a topological space X and a continuous scalar function $f : X \rightarrow \mathbb{R}$, the set $\{x \in X : f(x) = a\}$ is a *level set* of f for some value $a \in \mathbb{R}$. The level sets of f may have multiple connected components. The Reeb graph of f is obtained by continuously collapsing each connected component in the level set into a single point. Intuitively, as a changes continuously, the connected components in the level sets appear, disappear, split and merge; and the Reeb graph of f tracks such changes. Hence, the Reeb graph provides a simple yet meaningful abstraction of the input scalar field. It has been used in a range of applications in computer graphics and visualization; see, for example, the survey [2] and references therein on applications of Reeb graph.

Our results. Most of the previous work on the Reeb graph focused on its efficient computation. In this paper, we initiate the study of two questions related to Reeb graphs both of which are important in shape and data analysis applications.

The first question is concerned with the approximation of the Reeb graph from a set of points sampled from a hidden manifold. It turns out that the Reeb graph homology is also related to the so-called vertical homology groups. These relations enable us to develop an efficient algorithm to approximate the Reeb graph of the manifold from its point samples.

*Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43221. Emails: tamaldehy, yusu@cse.ohio-state.edu.

As a by-product of our approximation result, we also obtain a near-linear time algorithm that computes the first betti number $\beta_1(M)$ of an orientable smooth compact 2-manifold M without boundary from its point samples. This result may be of independent interest even though the correctness of our algorithm needs a slightly stronger condition than the previous best-known approach for computing $\beta_1(M)$ from point data. Using a result of Hausmann [18], one can compute the first betti number of a Rips complex constructed out of the input data and claim it as $\beta_1(M)$. A straightforward computation of betti numbers of the Rips complex using Smith normal form [19] takes cubic time whereas our algorithm runs in near-linear expected time.

The second question we study concerns with the definition and computation of loops in Reeb graphs which remain “persistent” as its defining domain “grows”. We propose a definition of the *persistent Reeb graph homology* for a sequence of Reeb graphs computed for a function defined on a filtered space in the same spirit as the standard persistent homology which is defined for a filtered space [16]. Interestingly, this problem does not seem to be easier than computing the standard persistent homology, potentially due to the fact that the domains in question (the sequence of Reeb graphs) do not have an inclusion between them, as was the case for standard persistence homology.

Related work. As mentioned already, most previous work on the Reeb graph focused on its efficient computation. Shinagawa and Kunii [22] presented the first provably correct algorithm to compute Reeb Graphs for a triangulation of a 2-manifold in $\Theta(m^2)$ time where m is the number of vertices in the triangulation. Cole-McLaughlin et al. [8] improved the running time to $O(m \log m)$. Doraiswamy and Natarajan [13] extended the sweeping idea to compute the Reeb graph in $O(n \log n (\log \log n)^3)$ time from a triangulation of a d -manifold, where n is the size of the 2-skeleton of the triangulation. For a piecewise-linear function defined on an arbitrary simplicial complex, a simple algorithm is proposed in [12] that runs in time $O(n \log n + L)$, where $L = \Theta(nm)$ is the total complexity of all level-sets passing through critical points. Tierny et al. [23] proposed an algorithm that computes the Reeb graph for a 3-manifold with boundary embedded in \mathbb{R}^3 in time $O(n \log n + hn)$, where h is number of independent loops in the Reeb graph. A streaming algorithm was presented in [21] to compute the Reeb graph for an arbitrary simplicial complex in an incremental manner in $\Theta(nm)$ time. Very recently, Harvey et al. [17] presented an efficient randomized algorithm to compute the Reeb graph for an arbitrary simplicial complex in $O(n \log m)$ expected running time. The Reeb graph for a time-varying function defined on a 3-dimensional space was studied in [15].

Recently a flurry of research has been initiated on estimating topological information from point data, such as computing ranks of homology groups [5], cut locus [10], and the shortest set of homology loops [11]. In [3], Chazal et al. initiated the study of approximating topological attributes of scalar functions from point data, and showed that the standard persistent diagram induced by a function can be approximated from input points. This result was later used in [4] to produce a clustering algorithm with theoretical guarantees. We remark that results from [3, 4] can be used to approximate the *loop-free* version of the Reeb graph (the so-called *contour tree*) from point data, thus providing a partial solution to our first question. However, it is unclear how to approximate loops in the Reeb graph which correspond to a *subset of essential loops* in the input domain which represent a subgroup of H_1 -homology.

2 Background and notations

Homology. A homology group of a topological space X encodes its topological connectivity. We consider the *simplicial* homology group if X is a simplicial complex, and consider the *singular* homology group otherwise, both denoted with $H_p(X)$ for the p th homology group. The definitions of these two homology groups can be obtained from any standard book on algebraic topology. Here we single out the concepts of p -chains and p -cycles in singular homology whose definitions are not as widely known in computational geometry as their simplicial counterparts. See [19] for detailed discussions on this topic.

A singular p -simplex for a topological space X is a map σ that maps the standard p -simplex $\Delta^p \subset \mathbb{R}^p$ continuously in X . A p -chain is a formal sum of singular p -simplices. A singular p -cycle in X is a p -chain whose boundary is a zero $(p - 1)$ -chain. Therefore, technically speaking, a p -chain or a p -cycle for X is a formal sum of maps. To access the geometry of X , for a p -chain $\alpha = \sigma_1 + \dots + \sigma_k$, we define $\cup_i \sigma_i(\Delta^p) \subseteq X$ as the *carrier* of α . Let a *loop* refer to the image of an injective map $\mathbb{S}^1 \rightarrow X$ or a finite union of such images. We will deal with 1-cycles whose carriers are loops.

We assume that X is triangulable. Thus, its simplicial homology defined by a triangulation identifies to its singular homology. We also assume that the homology groups are defined over \mathbb{Z}_2 coefficients. Since \mathbb{Z}_2 is a field, $H_p(X)$ is a vector space of dimension p .

Similar to homology groups, a cycle in X refers to a *simplicial* cycle when X is a simplicial complex and a *singular* cycle otherwise. Let $\mathbf{Z}_p(X)$ denote the p -th cycle group in X . A continuous map $\Phi : X_1 \rightarrow X_2$ between two topological spaces induces a map among its chain groups which we denote as $\Phi_\#$. Clearly, $\Phi_\#$ provides a map from the cycle group $\mathbf{Z}_p(X_1)$ to the cycle group $\mathbf{Z}_p(X_2)$ which in turn induces a homomorphism $\Phi_* : H_p(X_1) \rightarrow H_p(X_2)$.

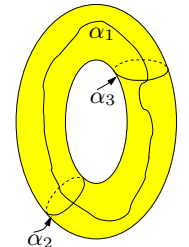
Horizontal and Vertical Homology Following [6], we now extend the standard homology to the so-called *horizontal* and *vertical homology with respect to a function* $f : X \rightarrow \mathbb{R}$. First, given a continuous function f , its *level sets* and *interval sets* are defined as: $X_a := f^{-1}(a)$ and $X_I := f^{-1}(I)$ for $a \in \mathbb{R}$ and for an open or closed interval $I \subseteq \mathbb{R}$, respectively. From now on we sometimes omit the use of f when its choice is clear from the context.

A homology class $\omega \in H_p(X)$ is *horizontal* if there exists a discrete set of iso-values $\{a_i\}$ such that ω has a pre-image under the map $H_p(\cup_i X_{a_i}) \rightarrow H_p(X)$ induced by inclusion. The set of horizontal homology classes form a subgroup $\overline{H}_p(X)$ of $H_p(X)$ since the trivial homology class is horizontal, and the addition of any two horizontal homology class is still horizontal. We call this subgroup $\overline{H}_p(X)$ the *horizontal homology group of X with respect to f* . The *vertical homology group of X with respect to f* is defined as:

$$\check{H}_p(X) := H_p(X)/\overline{H}_p(X), \text{ the quotient of } H_p(X) \text{ with } \overline{H}_p(X).$$

The coset $\omega + \overline{H}_p(X)$ for every class $\omega \in H_p(X)$ provides an equivalence class in $\check{H}_p(X)$. We call ω a *vertical homology class* if $\omega + \overline{H}_p(X)$ is not identity in $\check{H}_p(X)$. In other words, $\omega \notin \overline{H}_p(X)$. Two homology classes ω_1 and ω_2 are *vertically homologous* if $\omega_1 + \omega_2 \in \overline{H}_p(X)$.

We percolate the definitions from the homology classes to cycles. A cycle α is *horizontal* if $[\alpha]$, the standard homology class represented by α , is a horizontal class. Two cycles α_1 and α_2 are *vertically homologous* if $[\alpha_1]$ and $[\alpha_2]$ are vertically homologous. Obviously, two p -cycles α_1 and α_2 are vertically homologous if and only if there is a $(p + 1)$ -chain B such that $\partial B + \alpha_1 + \alpha_2$ is a horizontal cycle. See the torus in the right figure for an example, where α_2 is a horizontal cycle as it is homologous to α_3 carried by a loop contained in a connected component of a level set; while α_1 is a vertical cycle, i.e, $[\alpha_1]$ is a vertical homology class. We say that $\{\alpha_1, \dots, \alpha_k\}$ is a *set of base cycles for $H_p(X)$* if $\{[\alpha_1], \dots, [\alpha_k]\}$ form a basis for $H_p(X)$. A set of base cycles for $\overline{H}_p(X)$ and $\check{H}_p(X)$ are defined analogously.



Finally, the *range* of a loop $\gamma \subseteq X$, denoted by $\text{range}(\gamma)$, is the interval $[\min_{x \in \gamma} f(x), \max_{x \in \gamma} f(x)]$. The *height* of this loop, $\text{height}(\gamma)$, is simply the length of $\text{range}(\gamma)$. We extend the definitions of range and height to cycles by saying that $\text{range}(\alpha) = \text{range}(\gamma)$ and $\text{height}(\alpha) = \text{height}(\gamma)$ where the cycle $\alpha \in \mathbf{Z}_1(X)$ is carried by the loop γ . The *height of a homology class* ω , denoted by $\text{height}(\omega)$, is the minimal height of any cycle in this class. Notice that the height of a horizontal class ω is not necessarily zero since ω may be the addition of multiple height-0 horizontal classes.

Reeb graph. Given a triangulable topological space X and a continuous function $f : X \rightarrow \mathbb{R}$, we say that two points $x, y \in X$ are *equivalent*, denoted by $x \sim y$, if and only if x and y belong to the same connected

component of X_a for some $a \in \mathbb{R}$. Consider the quotient space X_{\sim} , which is the set of equivalence classes equipped with the quotient topology induced by this equivalence relation; X_{\sim} is also called the *Reeb graph* of X with respect to f , denoted by $R_f(X)$. See Figure 1 (a) and (b) for an example.

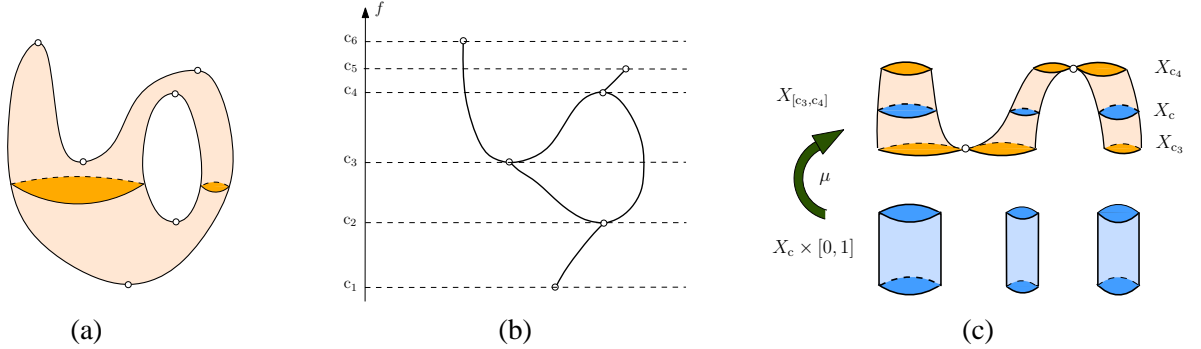


Figure 1: (a) X is a solid torus and its Reeb graph w.r.t the height function f is shown in (b). (c) If f is levelset-tame, then there is a continuous map $\mu : X_c \times [0, 1] \rightarrow X_{[c_3, c_4]}$ whose restriction to the open set $X_c \times (0, 1)$ is a homeomorphism.

An alternative way to view the Reeb graph is that there is a natural continuous surjection $\Phi : X \rightarrow X_{\sim}$ where $\Phi(x) = \Phi(y)$ if and only if x and y come from the same connected component of a level set of f . In this sense, $R_f(X)$ is obtained by continuously identifying each connected component. The map Φ induces a scalar function $\tilde{f} : R_f(X) \rightarrow \mathbb{R}$ where $\tilde{f}(p) = f(x)$ if $p = \Phi(x)$. Since $f(x) = f(y)$ whenever $\Phi(x) = \Phi(y)$, the function \tilde{f} is well-defined. Since f is continuous, so is \tilde{f} . The range or height of a loop in $R_f(X)$ is measured with respect to this function \tilde{f} . In this paper, we abuse the notation slightly and use f to also refer to \tilde{f} for simplicity.

3 Reeb graphs and vertical homology

In this section, we show that $H_1(R_f(X))$ and the first vertical homology group $\check{H}_1(X)$ of X are isomorphic¹.

The surjection $\Phi : X \rightarrow R_f(X)$ induces a chain map $\Phi_{\#}$ from the 1-chains of X to the 1-chains of $R_f(X)$ which eventually induces a homomorphism $\Phi_* : H_1(X) \rightarrow H_1(R_f(X))$. For the horizontal subgroup $\overline{H}_1(X)$, we have that $\Phi_*(\overline{H}_1(X)) = \emptyset = \overline{H}_1(R_f(X))$. Hence Φ_* induces a well-defined homomorphism between the quotient groups

$$\check{\Phi} : \check{H}_1(X) = \frac{H_1(X)}{\overline{H}_1(X)} \rightarrow \frac{H_1(R_f(X))}{\overline{H}_1(R_f(X))} = H_1(R_f(X)).$$

In what follows, we show that $\check{\Phi}$ is indeed an isomorphism under some mild conditions. Intuitively, this is not surprising as Φ maps each contour in the level set to a single point, which in turn also collapses every horizontal cycle.

For technical reasons, we consider functions that behave nicely. Specifically, we call a continuous function $f : X \rightarrow \mathbb{R}$ *levelset-tame* if there are only finite number of discrete values $\{c_1, \dots, c_m\}$ such that the following is true: for any two consecutive c_i and c_{i+1} , (i) there is a homeomorphism $\mu_i : X_c \times (0, 1) \rightarrow X_{(c_i, c_{i+1})}$ for an arbitrary $c \in (c_i, c_{i+1})$; (ii) the homeomorphism μ_i can be extended to a continuous map $\mu_i : X_c \times [0, 1] \rightarrow X_{[c_i, c_{i+1}]}$; (iii) there is no map satisfying (i) and (ii) whose extension in (ii) is also a homeomorphism. We call c_i s the *levelset-critical values*. See Figure 1 (c) for an example. It can be shown

¹This relation is observed for 2-manifolds in [6], but to the best of our knowledge, it has not been formally introduced and proved anywhere yet for general topological spaces. We include it here for completion.

that Morse functions on a compact smooth manifold and piecewise-linear functions on a finite simplicial complex are both levelset-tame functions.

The set of points from $R_f(X)$ having levelset-critical values are called the *nodes* of the Reeb graph $R_f(X)$. (Some nodes may have degree 2). Removing the nodes from $R_f(X)$ leaves a set of connected components. The closure of each such component is an *arc* of $R_f(X)$. Given an arc γ of $R_f(X)$, let $X(\gamma)$ denote the pre-images of γ under the map Φ . Observe that any two points of $X(\gamma)$ are path-connected within $X(\gamma)$.

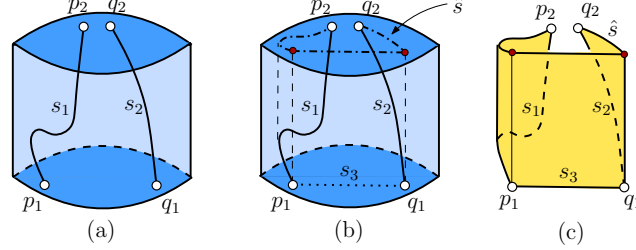


Figure 2: (a) An illustration of the cylinder $\mathcal{C} = S \times [0, 1]$, where each horizontal slice of this cylinder is a copy of S . (b) \hat{s} is the projection of $s = s_1 \circ s_3 \circ s_2$ from the product space onto the slice $\mathcal{C}[1]$. (c) The boundary of the surface B' is $s + \hat{s}$.

We now show two results to relate vertical cycles in X and cycles in $R_f(X)$ for a levelset-tame function f . Our main result of this section is obtained from these two claims.

Claim 3.1 *Let $f : X \rightarrow \mathbb{R}$ be a levelset-tame function. Given a loop $\gamma \subseteq R_f(X)$, there is a loop $\hat{\gamma} \subseteq X$ such that $\Phi(\hat{\gamma}) = \gamma$ and $\text{range}(\gamma) = \text{range}(\hat{\gamma})$.*

Proof: We construct $\hat{\gamma}$ from γ as follows: suppose γ consists of a sequences of k arcs

$$\gamma[p_1, p_2], \gamma[p_2, p_3], \dots, \gamma[p_k, p_1]$$

where each p_i is a node of the Reeb graph $R_f(X)$. For each p_i , choose an arbitrary pre-image $q_i \in X$ from $\Phi^{-1}(p_i)$. Now for each arc $\gamma[p_i, p_{i+1}]$, connect q_i and q_{i+1} within $X(\gamma[p_i, p_{i+1}])$ by $\hat{\gamma}[q_i, q_{i+1}]$ arbitrarily. The concatenation of all $\hat{\gamma}[q_i, q_{i+1}]$ provides $\hat{\gamma} \subset X$. It is easy to check that $\text{range}(\gamma) = \text{range}(\hat{\gamma})$. ■

Claim 3.2 *Let $f : X \rightarrow \mathbb{R}$ be a levelset-tame function. For a 1-cycle $\alpha \in \mathbf{Z}_1(X)$ the image $\Phi_{\#}(\alpha)$ represents a trivial class in $H_1(R_f(X))$ if and only if $[\alpha]$ is horizontal in $H_1(X)$.*

Proof: First, if a cycle α is from a horizontal class $[\alpha]$, then there is another representative cycle α' of $[\alpha]$ such that the carrier of α' is contained in a discrete set of level-sets. Since each connected component within a level-set is mapped to a single point in the Reeb graph, the image $\Phi_{\#}(\alpha')$ in $R_f(X)$ is a 0-chain and thus has trivial H_1 -homology. It follows that $\Phi_{\#}(\alpha)$ also has trivial homology.

We now show the opposite direction. That is, if the image of α in $R_f(X)$ has trivial homology, then $[\alpha]$ must be horizontal. For simplicity, assume that $\Phi_{\#}(\alpha)$ is carried by a sequence of full arcs in $R_f(X)$. The case where $\Phi_{\#}(\alpha)$ is carried by some arcs only partially can be handled similarly. Since $\Phi_{\#}(\alpha)$ has trivial homology in $H_1(R_f(X))$, each arc of $R_f(X)$ appears even number of times in its carrier. Consider an arc γ of $R_f(X)$ and its pre-image $X(\gamma)$ in X . Let c_p and c_{p+1} be the levelset-critical values of the endpoints of γ . Assume that the carrier of α consists of $2k$ pieces in $X(\gamma)$, whose chains are $\alpha_1, \alpha_2, \dots, \alpha_{2k}$. Partition it into k pairs $(\alpha_1, \alpha_2), \dots, (\alpha_{2k-1}, \alpha_{2k})$. Consider an arbitrary pair $(\alpha_{2i}, \alpha_{2i+1})$. We now argue that there exists a 2-chain B_i such that the carrier of $\partial B_i + \alpha_{2i} + \alpha_{2i+1}$ lies in the level sets $X_{c_p} \cup X_{c_{p+1}}$. Notice that by taking

the union of all such B_i s for $i \in [1, k]$, we obtain a 2-chain B_γ such that the carrier of $\partial B_\gamma + \alpha_1 + \dots + \alpha_{2k}$ is contained in the two level sets X_{c_p} and $X_{c_{p+1}}$. Taking the union of B_γ for all arcs γ from $R_f(X)$ gives rise to a 2-chain B such that the carrier of $\partial B + \alpha$ is contained in a discrete set of level sets. It follows that α is a horizontal cycle.

Now, we only need to focus on the construction of the 2-chain B_i for a pair of chains α_{2i} and α_{2i+1} . Let π_1 and π_2 be the two curves carrying α_{2i} and α_{2i+1} . Recall that there is a continuous map $\mu : X_c \times [0, 1] \rightarrow X_{[c_p, c_{p+1}]}$ for an arbitrary but fixed $c \in (c_p, c_{p+1})$, whose restriction to the open set $X_c \times (0, 1)$ is a homeomorphism onto $X_{(c_p, c_{p+1})}$. Let π_1° and π_2° denote the interiors of π_1 and π_2 , respectively; π_1° and π_2° have unique pre-images s_1° and s_2° in $X_c \times (0, 1)$ under μ .

The product space $X_c \times [0, 1]$ has several connected components each of which, called a *cylinder*, corresponds to the product between a connected component in the level-set X_c and $[0, 1]$. The images of all such cylinders under μ can touch each other only in X_{c_p} or in $X_{c_{p+1}}$ when μ is no longer a homeomorphism. See Figure 1 (c) for an illustration, where in this example, $X_c \times [0, 1]$ has three cylinders. The cylinder that contains s_1° and s_2° is denoted as $\mathcal{C} = S \times [0, 1]$, where S is the corresponding connected component in X_c . Every point $\mathbf{x} \in \mathcal{C}$ can be represented as $\mathbf{x} = (x, t)$, where $x \in S$ is called its *horizontal coordinate* and $t \in [0, 1]$ is its *vertical coordinate* (or *height*). A *slice* $\mathcal{C}[t]$ refers to one copy of S at height t .

Let s_1 (resp. s_2) denote the closure of s_1° (resp. s_2°) in \mathcal{C} , with p_1 and p_2 (resp. q_1 and q_2) being its endpoints. See Figure 2 (a) for an illustration. Notice that $\mu(s_1) = \pi_1$ and $\mu(s_2) = \pi_2$ due to the continuity of μ . Since each slice $\mathcal{C}[t]$ of the cylinder \mathcal{C} is path-connected, there is a path, say s_3 , that connects p_1 and q_1 in $\mathcal{C}[0]$. Let s denote the concatenated curve $s_1 \circ s_3 \circ s_2$; see Figure 2 (b). Now for every point $\mathbf{x} = (x, t_x) \in s$, consider the ‘‘vertical line’’ $l_x = \{(x, t) \mid t \in [t_x, 1]\}$. That is, l_x contains the images of \mathbf{x} in each slice $\mathcal{C}[t]$ with $t \geq t_x$. The union of l_x s for all $\mathbf{x} \in s$ traces out a 2-dimensional surface B' . The boundary of B' , denoted by $\text{bnd } B'$, is $\text{bnd } B' = s \circ \hat{s}$ where \hat{s} is the image of s in $\mathcal{C}[1]$. See Figure 2 (b) and (c).

Finally, through the continuous map μ , we obtain a 2-chain B_i whose carrier is $\mu(B') \subset X_{[c_p, c_{p+1}]}$ and $\text{bnd } \mu(B') = \pi_1 \circ \mu(s_3) \circ \pi_2 \circ \mu(\hat{s})$. Furthermore, $\mu(s_3) \cup \mu(\hat{s})$ lie in the level-sets $X_{c_p} \cup X_{c_{p+1}}$. As described earlier, by taking the union of such 2-chains for all pairs and for all arcs, we obtain a 2-chain B whose boundary is exactly α and a finite set of cycles whose carriers are contained in a discrete set of level-sets $\bigcup_{c_p} X_{c_p}$. Hence $[\alpha]$ must be a horizontal homology class. ■

Theorem 3.3 *Given a levelset-tame function $f : X \rightarrow \mathbb{R}$, let $\check{\Phi} : \check{H}_1(X) \rightarrow H_1(R_f(X))$ be the homomorphism induced by the surjection $\Phi : X \rightarrow R_f(X)$ as defined before. The map $\check{\Phi}$ is an isomorphism. Furthermore, for any vertical homology class $\omega \in \check{H}_1(X)$, we have that $\text{height}(\omega) = \text{height}(\check{\Phi}(\omega))$.*

Proof: It follows from Claim 3.1 that the homomorphism $\Phi_* : H_1(X) \rightarrow H_1(R_f(X))$ induced by Φ is surjective. Combining this with the fact that $\Phi_*(\overline{H}_1(X)) = \emptyset = \overline{H}_1(R_f(X))$ implies that the quotient map $\check{\Phi}$ is also surjective. The injectivity of $\check{\Phi}$ follows from Claim 3.2. Hence $\check{\Phi}$ is an isomorphism.

For the second part of the theorem, suppose α is a vertical cycle such that $[\alpha] = \omega$ and $\text{height}(\alpha) = \text{height}(\omega)$, i.e., α is a *thinnest cycle* in the vertical homology class ω . Let $\gamma \subseteq R_f(X)$ be the loop in $R_f(X)$ that carries the cycle in the homology class $\check{\Phi}(\omega) \in H_1(R_f(X))$. We have that

$$\text{height}(\alpha) \geq \text{height}(\Phi_\#(\alpha)) \geq \text{height}(\check{\Phi}(\omega)) = \text{height}(\gamma) \quad (1)$$

On the other hand, by Claim 3.1, there is a loop $\hat{\gamma} \subseteq X$ such that $\Phi(\hat{\gamma}) = \gamma$ and $\text{height}(\hat{\gamma}) = \text{height}(\gamma)$. Let $\hat{\alpha}$ be any 1-cycle carried by $\hat{\gamma}$. By Claim 3.2, we have $[\hat{\alpha}] = \omega$, as the cycle $\alpha + \hat{\alpha}$ is mapped to a trivial cycle in $R_f(X)$. Hence $\text{height}(\gamma) = \text{height}(\hat{\gamma}) \geq \text{height}(\alpha)$. Combining this with Eqn (1) proves that $\text{height}(\check{\Phi}(\omega)) = \text{height}(\omega)$. ■

4 Approximating Reeb graphs

Let M be a compact and smooth m -manifold without boundary embedded in \mathbb{R}^d . The *reach* $\rho(M)$ of M is the minimal distance from any point $x \in M$ to the so-called medial axis of M . Given a point $p \in M$, let $B_M(p, r)$ denote the open geodesic ball centered at p with radius r . Let r_p be the maximal radius so that $B_M(p, r_p)$ is convex in the sense that the minimizing geodesics between any two points in $B_M(p, r_p)$ is contained in $B_M(p, r_p)$. The *convexity radius* of M is simply $\rho_c(M) = \inf_{p \in M} r_p$.

A set of points P is an ε -*sample* of M if $P \subset M$ and for any point $x \in M$, there is a point $p \in P$ within ε geodesic distance from x . Given P and a real $r > 0$, the *Čech complex* $\mathcal{C}^r(P)$ is a simplicial complex where a simplex $\sigma \in \mathcal{C}^r(P)$ if and only if the vertices of σ are the centers of d -balls of radius $r/2$ with a non-empty common intersection. Instead of common intersection, if we only require pairwise intersection among the set of d -balls, we obtain the so-called *Vietoris-Rips complex* (*Rips complex for short*) $\mathcal{R}^r(P)$.

Overview. Consider an ε -sample $P \subset M$ and a function $f : M \rightarrow \mathbb{R}$ with its value only available at sample points in P . In what follows, we show that for an appropriate r , the Reeb graph of the Rips complex $\mathcal{R}^r(P)$ approximates $R_f(M)$ both in terms of the rank of the first homology group, and in terms of the range and the height of cycles and homology classes. Precise definition of approximation will be given later. Once the Rips complex is constructed, computing its Reeb graph takes only $O(n \log n)$ expected time [17], where n is the size of the 2-skeleton of $\mathcal{R}^r(P)$. Since f is only available at sample points in P , the approximation quality naturally depends on how well the function $f : M \rightarrow \mathbb{R}$ behaves. We assume that f is Lipschitz with Lipschitz constant Lip_f .

In Section 4.1 we first introduce some relations between cycles of M and those of the geometric realization $|\mathcal{R}^r(P)|$ of the Rips complex $\mathcal{R}^r(P)$. Using these relations, in Section 4.2, we show that there are maps between $H_1(M)$ and $H_1(|\mathcal{R}^r(P)|)$ that are not only isomorphic, but also preserves the height / range of a homology class. This, combined with Theorem 3.3, eventually leads to our approximation of $R_f(M)$. This approximation result can be used to estimate the first betti number of an orientable 2-manifold from its point samples in near-linear expected time.

4.1 Relation between cycles in M and $|\mathcal{R}^r(P)|$

The simplicial complex $\mathcal{R}^r(P)$ as defined is not necessarily embedded in \mathbb{R}^d . Consider the embedding $e : \mathcal{R}^r(P) \rightarrow \Delta^{|P|}$ of $\mathcal{R}^r(P)$ into the standard simplex in $\mathbb{R}^{|P|}$. Let $|\mathcal{R}^r(P)|$ denote the underlying space of the geometric realization $e(\mathcal{R}^r(P))$. A piecewise-linear function f on $\mathcal{R}^r(P)$ defines naturally a piecewise linear function on its geometric realization $|\mathcal{R}^r(P)|$ which we also denote as f . The Reeb graph of a PL-function f on $\mathcal{R}^r(P)$ is in fact the Reeb graph of f on its geometric realization $|\mathcal{R}^r(P)|$. Hence $R_f(\mathcal{R}^r(P)) := R_f(|\mathcal{R}^r(P)|)$. Analogously, the vertical / horizontal homology groups of $\mathcal{R}^r(P)$ with respect to a PL-function f are also defined using $|\mathcal{R}^r(P)|$. In this section, we relate cycles from M and those from $|\mathcal{R}^r(P)|$ via (simplicial) cycles of $\mathcal{R}^r(P)$. We will show how to construct the maps as indicated in Figure 3 below, such that these maps not only induce isomorphisms in the corresponding homology groups, but also preserve height and range of cycles.

A general version of the next claim which establishes an isomorphism between the homology groups of M and those of Čech and Rips-complexes is well-known (see, e.g. [20] for Čech-complexes and [18] for Rips-complexes; a stronger variant for compact spaces was also observed by Steve Oudot (personal communications)). We include a proof of it for completeness.

²Here ε -sample is not defined relative to reach or feature size as commonly done in reconstruction literature [9].

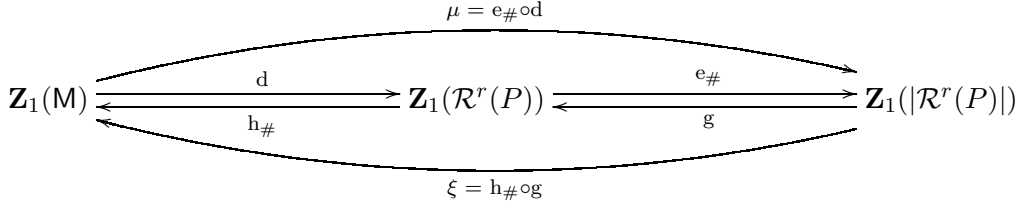


Figure 3: Maps between cycle groups

Claim 4.1 *Let $P \subset M$ be an ε -sample and r a parameter such that $4\varepsilon \leq r \leq \frac{1}{2}\sqrt{\frac{3}{5}}\rho(M)$. Then,*

$$H_1(\mathcal{C}^r(P)) \approx H_1(\mathcal{R}^r(P)) \approx H_1(\mathcal{C}^{2r}(P)) \approx H_1(M).$$

The first two isomorphisms are induced by the natural inclusion from $\mathcal{C}^r(P)$ to $\mathcal{R}^r(P)$ and then to $\mathcal{C}^{2r}(P)$. The last isomorphism is induced by the homotopy equivalence defined in Proposition 3.3 of [11].

Proof: Consider the following sequence of inclusions:

$$\mathcal{C}^r(P) \xrightarrow{i_1} \mathcal{R}^r(P) \xrightarrow{i_2} \mathcal{C}^{2r}(P) \xrightarrow{i_3} \mathcal{R}^{2r}(P) \xrightarrow{i_4} \mathcal{C}^{4r}(P).$$

Let $i_* : H_1(\mathcal{R}^r(P)) \rightarrow H_1(\mathcal{R}^{2r}(P))$ be induced by the inclusion $i = i_3 \circ i_2$. By Proposition 3.4 [11], we have that $image(i_*) \approx H_1(\mathcal{C}^r(P))$.

Notice that since $\mathcal{C}^{2r}(P)$ and $\mathcal{R}^{2r}(P)$ share the same edge set, and $\mathcal{R}^{2r}(P)$ only has more triangles than $\mathcal{C}^{2r}(P)$, the inclusion induces a surjection from $H_1(\mathcal{C}^{2r}(P))$ to $H_1(\mathcal{R}^{2r}(P))$. Combining this with the fact that $image(i_*) \approx H_1(\mathcal{C}^r(P)) \approx H_1(M) \approx H_1(\mathcal{C}^{2r}(P))$, one obtains that exactly $rank(H_1(\mathcal{C}^{2r}(P)))$ number of independent homology classes survive from $H_1(\mathcal{R}^r(P))$ to $H_1(\mathcal{R}^{2r}(P))$ through $H_1(\mathcal{C}^{2r}(P))$ via inclusions. Hence we have $H_1(\mathcal{C}^{2r}(P)) \approx H_1(\mathcal{R}^{2r}(P))$ where the isomorphism is induced by inclusion. By a similar argument we also have that $H_1(\mathcal{C}^r(P)) \approx H_1(\mathcal{R}^r(P))$ and the isomorphism is also induced by inclusion. Since $H_1(\mathcal{C}^r(P)) \approx H_1(\mathcal{C}^{2r}(P))$, we have $H_1(\mathcal{C}^r(P)) \approx H_1(\mathcal{R}^r(P)) \approx H_1(\mathcal{C}^{2r}(P)) \approx H_1(\mathcal{R}^{2r}(P))$, and all the isomorphisms are induced by inclusions. ■

Maps d and $h_{\#}$. We now define maps as indicated in Figure 3. First, given a cycle $\alpha \in \mathbf{Z}_1(M)$, we map it to a cycle $d(\alpha) \in \mathbf{Z}_1(\mathcal{R}^r(P))$ using the same Decomposition method [1] as applied in [11]. In particular, break the carrier of α into pieces where each piece has length at most $r - 2\varepsilon$. For each piece with endpoints x_i and x_{i+1} , find the closest sample points p_i and p_{i+1} from P to x_i and x_{i+1} , respectively, and connect p_i and p_{i+1} (which is necessarily an edge in $\mathcal{R}^r(P)$ by triangle inequality). The resulting simplicial 1-cycle in $\mathcal{R}^r(P)$ is $d(\alpha)$.

We now define the map $h : \mathcal{R}^r(P) \rightarrow M$ as the inclusion map $\mathcal{R}^r(P) \hookrightarrow \mathcal{C}^{2r}(P)$ composed with the homotopy equivalence from $\mathcal{C}^{2r}(P)$ to M introduced in Proposition 3.3 of [11]. The corresponding chain map $h_{\#}$ maps p -cycles to p -cycles, and p -boundaries to p -boundaries inducing the homomorphism $h_* : H_p(\mathcal{R}^r(P)) \rightarrow H_p(M)$. We restrict h_* only to the first homology group $h_* : H_1(\mathcal{R}^r(P)) \rightarrow H_1(M)$. By Claim 4.1, h_* is an isomorphism.

The following lemma states that $d : \mathbf{Z}_1(M) \rightarrow \mathbf{Z}_1(\mathcal{R}^r(P))$ is in fact the homology-inverse of $h_{\#}$, which eventually implies that the map d takes homologous cycles to homologous cycles and induces a well-defined homomorphism (in fact, an isomorphism) $d_* : H_1(M) \rightarrow H_1(\mathcal{R}^r(P))$. The ranges of mapped cycles are also related. We put the proof of the following lemma in Appendix A to maintain the flow of the presentation. Given two intervals $I_1 = [a, b]$ and $I_2 = [c, d]$, we say that I_1 is *onside- δ -close* to I_2 if: $[a, b] \subseteq [c - \delta, d + \delta]$. I_1 and I_2 are *δ -Hausdorff-close* if the two intervals are onside- δ -close to each other. In the Lemma below,

assume that f is a (Lip_f) -Lipschitz function on M and its values for the vertices $P \subset M$ defines a piecewise linear function on $\mathcal{R}^r(P)$ which we also denote as f .

Lemma 4.2 (i) *The map d induces a well-defined homomorphism $d_*: H_1(M) \rightarrow H_1(\mathcal{R}^r(P))$. In fact, $h_* = (d_*)^{-1}$; that is, $h_* \circ d_* = id_{H_1(M)}$ and $d_* \circ h_* = id_{H_1(\mathcal{R}^r(P))}$.*

(ii) *Both $d_*: H_1(M) \rightarrow H_1(\mathcal{R}^r(P))$ and $h_*: H_1(\mathcal{R}^r(P)) \rightarrow H_1(M)$ are isomorphisms.*

(iii) *The range of the cycle $d(\alpha) \in \mathbf{Z}_1(\mathcal{R}^r(P))$ is onside- $(r \cdot \text{Lip}_f)$ -close to the range of $\alpha \in \mathbf{Z}_1(M)$. Similarly, the range of the cycle $h_{\#}(\hat{\alpha}) \in \mathbf{Z}_1(M)$ is onside- $(\cdot \text{Lip}_f)$ -close to $\hat{\alpha} \in \mathbf{Z}_1(\mathcal{R}^r(P))$.*

(iv) *The ranges of any homology class $\omega \in H_1(M)$ (resp. $\hat{\omega} \in H_1(\mathcal{R}^r(P))$) and its image $d_*(\omega) \in H_1(\mathcal{R}^r(P))$ (resp. $h_*(\hat{\omega}) \in H_1(M)$), are $(r \cdot \text{Lip}_f)$ -Hausdorff-close.*

Maps $e_{\#}$ and g . The map e between $\mathcal{R}^r(P)$ and its geometric realization $|\mathcal{R}^r(P)|$ is quite standard. It embeds $\mathcal{R}^r(P)$ into the standard simplex $\Delta \subset \mathbb{R}^{|P|}$ as follows: Each vertex $p_i \in P$ is mapped to the point $v_i = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^{|P|}$ with the i th position 1; and a simplex in $\mathcal{R}^r(P)$ with vertices $\{p_{i_0}, \dots, p_{i_r}\}$ is mapped to the simplex in $\mathbb{R}^{|P|}$ with vertices $\{v_{i_0}, \dots, v_{i_r}\}$. The chain map $e_{\#}$ induced by e provides the necessary map between $\mathbf{Z}_1(\mathcal{R}^r(P))$ and $\mathbf{Z}_1(|\mathcal{R}^r(P)|)$. Clearly, this map preserves the range of a cycle.

We now define the map $g: \mathbf{Z}_1(|\mathcal{R}^r(P)|) \rightarrow \mathbf{Z}_1(\mathcal{R}^r(P))$ as follows: Consider a cycle α in $|\mathcal{R}^r(P)|$. The carrier of α passes through a sequence of simplices S : if a point in the carrier is contained in multiple simplices of $|\mathcal{R}^r(P)|$, then keep the one with the minimum dimension. Let $S = \{\sigma_1, \dots, \sigma_m\}$. Now choose an arbitrary but fixed vertex u_i for each σ_i , and let $p_{u_i} \in P$ denote the unique pre-image of u_i in $\mathcal{R}^r(P)$ under the embedding map e . Notice that for any two consecutive simplices σ_i and σ_{i+1} that the carrier of α passes through, it is necessary that either σ_i is face of σ_{i+1} or σ_{i+1} is a face of σ_i . Hence either $p_{u_i} = p_{u_{i+1}}$ or $p_{u_i}p_{u_{i+1}}$ is an edge in $\mathcal{R}^r(P)$. Therefore, we map α simply to the cycle $g(\alpha)$ given by the sequence of vertices $(p_{u_1}, \dots, p_{u_m}, p_{u_1})$ and edges between them. We have the following result about maps $e_{\#}$ and g .

Lemma 4.3 (i) *Every cycle α in $\mathcal{R}^r(P)$ is mapped to a cycle $e_{\#}(\alpha)$ with the same range in $|\mathcal{R}^r(P)|$ under the embedding $e: \mathcal{R}^r(P) \rightarrow |\mathcal{R}^r(P)|$. The map $e_*: H_1(\mathcal{R}^r(P)) \rightarrow H_1(|\mathcal{R}^r(P)|)$ is an isomorphism, and the ranges of any homology class $\omega \in H_1(\mathcal{R}^r(P))$ and its image $e_*(\omega) \in H_1(|\mathcal{R}^r(P)|)$ are also the same.*

(ii) *Every cycle α in $|\mathcal{R}^r(P)|$ is mapped to a cycle $g(\alpha)$ in $\mathcal{R}^r(P)$ whose range is onside- $(r \cdot \text{Lip}_f)$ -close to that of α . The map $g: \mathbf{Z}_1(|\mathcal{R}^r(P)|) \rightarrow \mathbf{Z}_1(\mathcal{R}^r(P))$ induces an isomorphism $g_*: H_1(|\mathcal{R}^r(P)|) \rightarrow H_1(\mathcal{R}^r(P))$, and $g_* = (e_*)^{-1}$. The ranges of any homology class $\hat{\omega} \in H_1(|\mathcal{R}^r(P)|)$ and its image $g_*(\hat{\omega}) \in H_1(\mathcal{R}^r(P))$ are $(r \cdot \text{Lip}_f)$ -Hausdorff close.*

Proof: Part (i) of the Lemma holds easily from the definition of the natural embedding $e: \mathcal{R}^r(P) \rightarrow |\mathcal{R}^r(P)|$.

For the second part, first observe that for any cycle α from $|\mathcal{R}^r(P)|$, we have that $[e_{\#} \circ g(\alpha)] = [\alpha]$. Indeed, by the construction of g , it is easy to verify that $e_{\#} \circ g(\alpha)$ and α are homotopic. Since e induces an isomorphism from $\mathcal{R}^r(P)$ to $|\mathcal{R}^r(P)|$, it follows that g maps homologous cycles in $|\mathcal{R}^r(P)|$ to homologous cycles in $\mathcal{R}^r(P)$. Hence g induces a well-defined map $g_*: H_1(|\mathcal{R}^r(P)|) \rightarrow H_1(\mathcal{R}^r(P))$. Furthermore, $g \circ e_{\#}(\alpha') = \alpha'$ for any cycle $\alpha' \in \mathcal{R}^r(P)$. It follows that g_* is the inverse of e_* and hence is an isomorphism.

Finally, note that for each simplex $\sigma \in |\mathcal{R}^r(P)|$, the function value difference between any two points $x, y \in \sigma$ is bounded by $r \cdot \text{Lip}_f$. Let γ be the carrier of a cycle α in $|\mathcal{R}^r(P)|$. By the construction of g , for each piece $\gamma \cap \sigma_i$ of γ within the simplex $\sigma_i \in S$, we have that $|f(x) - f(u_i)| \leq r \cdot \text{Lip}_f$ for any point $x \in \gamma \cap \sigma_i$. Since $f(u_i) = f(p_{u_i})$, we have:

$$[\min_{i \in [1, m]} f(p_{u_i}) + r \cdot \text{Lip}_f, \max_{i \in [1, m]} f(p_{u_i}) - r \cdot \text{Lip}_f] \subseteq \text{range}(\alpha) \subseteq [\min_{i \in [1, m]} f(p_{u_i}) - r \cdot \text{Lip}_f, \max_{i \in [1, m]} f(p_{u_i}) + r \cdot \text{Lip}_f].$$

On the other hand, we have that $\text{range}(g(\alpha)) \subseteq [\min_{i \in [1, m]} f(p_{u_i}), \max_{i \in [1, m]} f(p_{u_i})]$. Hence $\text{range}(g(\alpha))$ is $\text{oneside-}(r \cdot \text{Lip}_f)$ -close to $\text{range}(\alpha)$. By a similar argument as in the proof of Lemma 4.2 (iv), the closeness between the corresponding homology classes follows. \blacksquare

Combining Lemma 4.2 and 4.3, we obtain a similar result for maps between $\mathbf{Z}_1(M)$ and $\mathbf{Z}_1(|\mathcal{R}^r(P)|)$.

Theorem 4.4 *There is a map $\mu := e_{\#} \circ d$ from $\mathbf{Z}_1(M)$ to $\mathbf{Z}_1(|\mathcal{R}^r(P)|)$ that induces an isomorphism $\mu_* : H_1(M) \rightarrow H_1(|\mathcal{R}^r(P)|)$. The range of cycle $\mu(\alpha)$ is $\text{oneside-}(r \cdot \text{Lip}_f)$ -close to the range of α .*

There is a map $\xi := h_{\#} \circ g$ from $\mathbf{Z}_1(|\mathcal{R}^r(P)|)$ to $\mathbf{Z}_1(M)$ that induces an isomorphism $\xi_ : H_1(|\mathcal{R}^r(P)|) \rightarrow H_1(M)$. The range of cycle $\xi(\hat{\alpha})$ is $\text{oneside-}(2r \cdot \text{Lip}_f)$ -close to the range of cycle $\hat{\alpha}$.*

Furthermore, μ_ is the inverse of ξ_* . That is, $\xi_* \circ \mu_* = \text{id}_{H_1(M)}$, and $\mu_* \circ \xi_* = \text{id}_{H_1(|\mathcal{R}^r(P)|)}$. Given any homology class $\omega \in H_1(M)$ (resp. $\hat{\omega} \in H_1(|\mathcal{R}^r(P)|)$) and its image $\mu_*(\omega) \in H_1(|\mathcal{R}^r(P)|)$ (resp. $\xi_*(\hat{\omega}) \in H_1(M)$), their ranges are $(2r \cdot \text{Lip}_f)$ -Hausdorff-close.*

4.2 $R_f(M)$ and $R_f(\mathcal{R}^r(P))$

We now show that under mild conditions on M , the induced isomorphisms μ_* and ξ_* as defined above in fact map horizontal classes to horizontal classes, and vertical classes to vertical classes.

Set $s = \text{rank}(\overline{H}_1(M))$. It follows from [6] that there is a set of base cycles $\{\alpha_1, \dots, \alpha_s\}$ for the horizontal subgroup $\overline{H}_1(M) \subseteq H_1(M)$ such that $\text{height}(\alpha_i) = 0$ for any $i \in [1, s]$. We call this set *0-height base cycles* for $\overline{H}_1(M)$. The corresponding set of homology classes $\{[\alpha_1], \dots, [\alpha_s]\}$ is called a *0-height basis* for $\overline{H}_1(M)$. For a horizontal homology class ω with height 0, the *span* of ω is the length of the maximal interval I such that ω has a pre-image in the levelset X_a for any $a \in I$. Intuitively, this is the interval in function values in which this homology class survives in the level-sets.

Let $s^*(M)$ denote the smallest span of any 0-height horizontal class of the input manifold M , and $t^*(M)$ the minimal height of any vertical class of M . We assume that both $s^*(M)$ and $t^*(M)$ are positive for our input level-set tame function on M .

Theorem 4.5 *Given a level-set tame function f on a manifold M , let $r > 0$ be such that $s^*(M), t^*(M) > 2r \cdot \text{Lip}_f$. Let μ_* and ξ_* be as defined in Theorem 4.4. Then we have that $\mu_*(\overline{H}_1(M)) = \overline{H}_1(|\mathcal{R}^r(P)|)$, $\xi_*(\overline{H}_1(|\mathcal{R}^r(P)|)) = \overline{H}_1(M)$ and $\check{H}_1(M) \approx \check{H}_1(|\mathcal{R}^r(P)|)$.*

Proof: For simplicity, in this proof let R denote $|\mathcal{R}^r(P)|$. Below we first show that $\mu_*(\overline{H}_1(M)) = \overline{H}_1(R)$. Consider a set of 0-height base cycles $\{\alpha_1, \dots, \alpha_s\}$ for $\overline{H}_1(M)$ with $s = \text{rank}(\overline{H}_1(M))$.

Take an arbitrary α_i for $i \in [1, s]$, and let $[a, b]$ denote the maximal interval³ such that $[\alpha_i]$ has a preimage in the levelset M_c for any $c \in [a, b]$. The span of $[\alpha_i]$ is $b - a$ and is at least $s^*(M) > 2r \cdot \text{Lip}_f$. Take a representative cycle γ_a from M_a and γ_b from M_b of the homology class $[\alpha_i]$. Set $I_a := [a - r \cdot \text{Lip}_f, a + r \cdot \text{Lip}_f]$ and $I_b := [b - r \cdot \text{Lip}_f, b + r \cdot \text{Lip}_f]$. It follows from Theorem 4.4 that the carrier of $\mu(\gamma_a)$ is contained in the interval levelset R_{I_a} while the carrier of $\mu(\gamma_b)$ is contained in R_{I_b} . ($[\mu(\alpha_i)] = [\mu(\gamma_a)] = [\mu(\gamma_b)]$ is a non-trivial homology class in $H_1(R)$.) Since $b - a > 2r \cdot \text{Lip}_f$, we have $I_a \cap I_b = \emptyset$. A simple application of the Mayer-Vietoris sequence provides that the homology class $[\mu(\alpha_i)]$ has a preimage in the levelset R_c for any $c \in [a + r \cdot \text{Lip}_f, b - r \cdot \text{Lip}_f]$, which in turn implies that $[\mu(\alpha_i)]$ is horizontal. (A similar argument is used in [6].)

Indeed, for any $c \in [a + r \cdot \text{Lip}_f, b - r \cdot \text{Lip}_f]$, consider the space $R_1 = R_{(-\infty, c]}$ and $R_2 = R_{[c, +\infty)}$. Notice that the carriers of $\mu(\gamma_a)$ and $\mu(\gamma_b)$ are contained in R_1 and R_2 respectively, $R_1 \cup R_2 = R$ and $R_1 \cap R_2 = R_c$. Consider the following subsequence of the Mayer-Vietoris sequence:

$$H_1(R_1 \cap R_2) \rightarrow H_1(R_1) \oplus H_1(R_2) \longrightarrow H_1(R_1 \cup R_2).$$

³Such maximal interval can be open. We assume it is closed for simplicity. The case when it is open can be handled similarly.

Since the homology class $[\mu(\alpha_i)]$ has pre-images both in $H_1(R_1)$ and $H_1(R_2)$, by the exactness of this sequence, it must have pre-image in $H_1(R_1 \cap R_2) = H_1(R_c)$. Hence $[\mu(\alpha_i)]$ is horizontal for any $i \in [1, s]$. This means that $\mu_*(\overline{H}_1(M))$ is a subgroup of $\overline{H}_1(R)$ and $s = \text{rank}(\overline{H}_1(M)) \leq \text{rank}(\overline{H}_1(R))$.

We now show that $t := \text{rank}(\check{H}_1(R)) \leq s$, which would imply that $\mu_*(\check{H}_1(M)) = \check{H}_1(R)$. Specifically, take a set of 0-height base cycles $\{\beta_1, \dots, \beta_t\}$ for $\check{H}_1(R)$. By Theorem 4.4, their images $\{\xi(\beta_1), \dots, \xi(\beta_t)\}$ in M is a set of independent cycles such that $\text{height}(\xi(\beta_i)) \leq 2r \cdot \text{Lip}_f$. Since the minimal height of any vertical cycle in M is $t^*(M) > 2r \cdot \text{Lip}_f$, each $\xi(\beta_i)$ has to be a horizontal homology cycle. It follows that $t \leq \text{rank}(\check{H}_1(M))$ as we wanted to show.

What remains to show is that $\check{H}_1(M) \approx \check{H}_1(R)$. Recall that μ_* is an isomorphism. Observe that

$$\overline{H}_1(M) \oplus \check{H}_1(M) \approx H_1(M) \approx \mu_*(H_1(M)) = H_1(R) \approx \overline{H}_1(R) \oplus \check{H}_1(R)$$

and we have already proved $\mu_*(\overline{H}_1(M)) = \overline{H}_1(R)$. It follows that $\check{H}_1(M) \approx \check{H}_1(R)$. ■

4.3 Putting everything together

We say that a Reeb graph $R_f(A)$ δ -approximates another Reeb graph $R_g(B)$ if (i) $H_1(R_f(A))$ is isomorphic to $H_1(R_g(B))$; and (ii) there is a one-to-one correspondence between cycles from $R_f(A)$ and $R_g(B)$ such that the ranges of each corresponding pair of cycles are δ -Hausdorff-close. Since a Reeb graph is a graph, each H_1 -homology class has only one representative cycle under \mathbb{Z}_2 coefficients. Hence a one-to-one correspondence between cycles in two Reeb graphs is equivalent to a one-to-one correspondence between their H_1 -homology classes. Combining Theorems 3.3, 4.4 and 4.5, we have our first main result.

Theorem 4.6 *Let $f : M \rightarrow \mathbb{R}$ be a level-set tame function defined on M with Lipschitz constant Lip_f . Given an ε -sample P of M , let r be a parameter such that $4\varepsilon \leq r < \min\{\frac{1}{4}\rho(M), \frac{1}{4}\rho_c(M), \frac{t^*}{2\text{Lip}_f}, \frac{s^*}{2\text{Lip}_f}\}$, and $\mathcal{R}^r(P)$ the Rips complex constructed from P using radius $r/2$. Then $R_f(\mathcal{R}^r(P))$ is a $(2r \cdot \text{Lip}_f)$ -approximation of $R_f(M)$, and $R_f(\mathcal{R}^r(P))$ can be computed in $O(n \log n)$ expected time [17], where n is the size of the 2-skeleton of $\mathcal{R}^r(P)$.*

Remark 1. Here we provide a brief discussion of why we focus only on the 1-st homology information of the Reeb graph, as well as the intuition behind our definition of a δ -approximate Reeb graph.

The Reeb graph is an abstract graph and contains only the 0- and 1-dimensional topological information. Given a Reeb graph $R_f(M)$, its 0-th homology simply encodes the connected components information of M , and can be approximated from point data easily by returning the number of connected components in an appropriately constructed Rips complex in linear time.

At the same time, compared to general abstract graphs, the Reeb graph has the extra information of the natural function f defined on it. Hence one may also ask what the 0-th *persistent* homology of $R_f(X)$ induced by f is. This turns out to be the same as approximating the 0-th persistent homology for X and can be solved using results from [3, 4].

Therefore, the only remaining issue is to approximate the 1-st homology of a Reeb graph. Similar to the case for the 0-th homology, there are two aspects: (i) computing $H_1(R_f(M))$ itself; and (ii) computing the 1-st *persistent* homology of $R_f(M)$ induced by the function f . For (i), our result shows that $H_1(R_f(\mathcal{R}^r(P)))$ for a certain Rips complex $\mathcal{R}^r(P)$ constructed from the point samples P is isomorphic to $H_1(R_f(M))$. For (ii), since every 1-cycle in a Reeb graph is essential, the standard persistence is not able to describe them, and one has to use the extended persistence as introduced in [6], which is determined by the range of essential cycles. Hence our definition of the approximation also requires that each cycle in one Reeb graph is mapped

uniquely to a cycle in the other one such that the ranges of these two cycles are close.

Remark 2. One can strengthen Theorem 4.6 slightly to show that if the parameter r does not satisfy the conditions that $r < \frac{\mathbf{t}^*}{2\text{Lip}_f}$ or $r < \frac{\mathbf{s}^*}{2\text{Lip}_f}$, then all homology classes of $H_1(R_f(M))$ with height at least $2r \cdot \text{Lip}_f$ are preserved in $H_1(R_f(\mathcal{R}^r(P)))$ (and vice versa).

Computing $\beta_1(M)$ for orientable 2-manifolds. It was shown in [8] that for a Morse function $f : M \rightarrow \mathbb{R}$ defined on a compact orientable surface M without boundary, one has $\text{rank}(H_1(M)) = 2 \cdot \text{rank}(H_1(R_f(M)))$. Hence intuitively, using Theorem 4.6, we can compute $\beta_1(M) = \text{rank}(H_1(M))$ by $2 \cdot \text{rank}(H_1(R_f(\mathcal{R}^r(P))))$ from an appropriate f and a Rips Complex $\mathcal{R}^r(P)$ constructed from a point sample P of M . Specifically, choose a function $f : M \rightarrow \mathbb{R}$ so that we can evaluate it at points in P . For example, pick a base point $\mathbf{v} \in P$ and define a function $f_{\mathbf{v}}(x)$ to be the Euclidean distance from $x \in M$ to the base point \mathbf{v} . Observe that the Lipschitz constant of this function $f_{\mathbf{v}}$ is at most 1. Our algorithm simply computes the Reeb graph $R_{f_{\mathbf{v}}}(\mathcal{R}^r(P))$ and returns $2 \cdot \text{rank}(H_1(R_{f_{\mathbf{v}}}(\mathcal{R}^r(P))))$.

Corollary 4.7 *Let M be an orientable smooth compact 2-manifold M without boundary and P an ε -sample of M . The above algorithm computes $\beta_1(M)$ in $O(n \log n)$ expected time if $\mathbf{t}^*(M)$ and $\mathbf{s}^*(M)$ are positive for the chosen function f , and the parameters satisfy $4\varepsilon \leq r < \min\{\frac{1}{4}\rho(M), \frac{1}{4}\rho_c(M), \frac{\mathbf{t}^*}{2\text{Lip}_f}, \frac{\mathbf{s}^*}{2\text{Lip}_f}\}$.*

Observe that a Morse function on an orientable 2-manifold provides positive \mathbf{t}^* and \mathbf{s}^* . We remark that our algorithm produces a correct answer only under good choices of f and r ; while previously, the best algorithm to estimate $\beta_1(M)$ only depends on choosing r small enough. The advantage of our algorithm is its efficiency, as the previous algorithm needs to compute the first-betti number of the simplicial complex $\mathcal{R}^r(P)$, which takes $O(n^3)$ time no matter what the intrinsic dimension of M is.

5 Persistent Reeb graph

Imagine that we have a set of points P sampled from a hidden space X , and $f : X \rightarrow \mathbb{R}$ a function whose values at points in P are available. We wish to study this function f through its Reeb graph. A natural approach to approximate X from P is to construct a Rips complex $\mathcal{R}^r(P)$ from P . Since it is often unclear what the right value of r should be, it is desirable to compute a series of Reeb graphs from Rips complexes constructed with various r , and then find out which cycles in the Reeb graph persist. This calls for computing *persistent homology groups* for the sequence of Reeb graphs.

Let $K_1 \subseteq K_2 \subseteq \dots \subseteq K_n$ be a filtration of a simplicial complex K_n . A piecewise linear function $f : |K_n| \rightarrow \mathbb{R}$ provides a PL-function for every K_i , $i \in [1, n]$. Let $R_f(K_i)$ denote the Reeb graph of f defined on the geometric realization $|K_i|$ of K_i and let $R_i := R_f(K_i)$. Below we first show that there is a sequence of homomorphisms $H_1(R_i) \rightarrow H_1(R_{i+1})$ induced by the inclusions $K_i \subset K_{i+1}$. We then present an algorithm to compute the persistent homologies induced by these homomorphisms.

5.1 Persistent Reeb graph homology

Let Φ_i denote the surjection from $|K_i| \rightarrow R_i$, for any $i \in [1, n]$. The maps Φ_i s, along with inclusions between K_j s, induce a well-defined continuous map $\xi : R_i \rightarrow R_j$, for any $i < j$. Indeed, given a point $p \in R_i$, points in its pre-image $\Phi_i^{-1}(p)$ come from the same contour in $|K_i|$ and share the same function value. Under inclusion $|K_i| \rightarrow |K_j|$ for any $j > i$, points in $\Phi_i^{-1}(p)$ are still contained in a single contour in $|K_j|$. Thus all points in the pre-image $\Phi_i^{-1}(p)$ are mapped to a single point $\xi(p) \in R_j$, implying that ξ is well-defined. Let ι_i denote the inclusion map from $|K_i|$ to $|K_j|$, and ξ_i the induced map from R_i to R_{i+1} . We have the following diagram that commutes.

$$\begin{array}{ccccccc}
|K_1| & \xrightarrow{\iota_1} & |K_2| & \xrightarrow{\iota_2} & \cdots & \xrightarrow{\iota_{n-1}} & |K_n| \\
\downarrow \Phi_1 & & \downarrow \Phi_2 & & & & \downarrow \Phi_n \\
R_1 & \xrightarrow{\xi_1} & R_2 & \xrightarrow{\xi_2} & \cdots & \xrightarrow{\xi_{n-1}} & R_n
\end{array}$$

The sequence of continuous maps ξ_i induces the following sequence of homomorphisms:

$$H_1(R_1) \xrightarrow{\xi_{1*}} H_1(R_2) \xrightarrow{\xi_{2*}} \cdots \xrightarrow{\xi_{(n-1)*}} H_1(R_n)$$

Following [14], we can now define the *persistent homology groups* as the images of maps $\xi_*^{i,j} = \xi_{j*} \circ \cdots \circ \xi_{i*} : H_1(R_i) \rightarrow H_1(R_j)$. In other words, the image $\text{Im}(\xi_*^{i,j})$ consists of homology classes from $H_1(R_j)$ that also have pre-images in $H_1(R_i)$ (i.e, persist from $H_1(R_i)$ to $H_1(R_j)$). The *persistent betti numbers* $\beta^{i,j}$ is defined as the rank of the persistent homology group $\text{Im}(\xi_*^{i,j})$. Set

$$\mu^{i,j} := \beta^{i-1,j} - \beta^{i,j} + \beta^{i,j-1} - \beta^{i-1,j-1}.$$

Intuitively, $\mu^{i,j}$ is the number of independent loops created upon entering R_i and destroyed upon leaving R_j . A *persistence pair* (i, j) is recorded if $\mu^{i,j} > 0$, and the value $\mu^{i,j}$ indicates the multiplicity of this pairing.

We focus on persistent H_1 -homology for R_i s in this paper. The persistent H_0 -homology for R_i s is the same as persistent H_0 -homology for K_i s, and thus can be easily computed by a union-find data structure in near linear time. We also remark that by Theorem 3.3, persistent H_1 -homology for R_i is isomorphic to persistent *vertical* homology $\check{H}(|K_i|)$ ⁴.

5.2 Computation

We now present an algorithm to compute the persistent betti numbers $\beta^{i,j}$. The numbers $\mu^{i,j}$ and the persistence pairs can be computed easily once we have these persistence betti numbers.

Given a filtration $K_1 \subseteq \cdots \subseteq K_n$, assume $K_{i+1} \setminus K_i$ is one simplex. Since the Reeb graph is completely decided by the 2-skeleton of a simplicial complex, we assume that K_i s are 2-complexes. Let n_v , n_e and n_t denote the number of vertices, edges and triangles in K_n , and $n = n_v + n_e + n_t$. Observe that the complexity of each Reeb graph R_i , for $i \in [1, n]$, is bounded by $O(n_e)$. The set of Reeb graphs R_i s can be computed in $O(nn_v)$ time using the incremental algorithm from [21]. We use this algorithm as it can also maintain the image of each edge from K_i in R_i in $O(n_v)$ time at each incremental step, thus providing Φ_i , for $i \in [1, n]$.

Recall that a set of base cycles for $H_p(\cdot)$ is a set of cycles whose classes form a basis of $H_p(\cdot)$. For the sake of exposition in this section, we abuse the notation slightly and use a cycle to also refer to its carrier in the Reeb graph. Specifically, we will see later that our algorithm in fact maintains the carriers of a set of base cycles for $H_1(R_i)$, which we also call a *cycle-basis*. We say that a set of cycles are *independent* if the set of homology classes these cycles represent are independent.

To compute $\beta^{i,j}$, one can construct a set of base cycles $\{\alpha_1, \dots, \alpha_r\}$ for $H_1(R_i)$ with $r = \text{rank}(H_1(R_i))$, and check how many of their images in R_j remain independent. A straightforward implementation of this approach takes $O(n^2 n_e^3)$ time. Indeed, $r = O(n_e)$ and the complexity of each cycle α_i is bounded by $O(n_v)$ (by representing them as a sequence of vertices). Computing the images of all α_i s takes $O(r n_v^2) = O(n_e n_v^2)$ time using the incremental algorithm from [21], and the independence test for these r cycles take $O(r n_e^2) = O(n_e^3)$ time. Finally, there are n^2 pairs of i and j that we need to test, giving rise to $O(n^2 n_e^3)$ total time complexity. To improve the time complexity, we follow the idea of the standard persistence algorithm [16] and perform only one scan of the sequence of Reeb graphs, while maintaining a set of base cycles at

⁴ Apriori, it is not clear how to compute the persistent one dimensional vertical homology. By maintaining the *extended persistence pairings* dynamically as we change from K_i to K_{i+1} , we can maintain the rank of each $H_1(K_i)$, but not the persistence homology between them.

any moment during the course. Notice that the standard persistence algorithm cannot be directly applied to the sequence of Reeb graphs as there are no inclusions among them. In fact, the underlying spaces of two consecutive Reeb graphs can change dramatically. See Figure 4 for such an example.

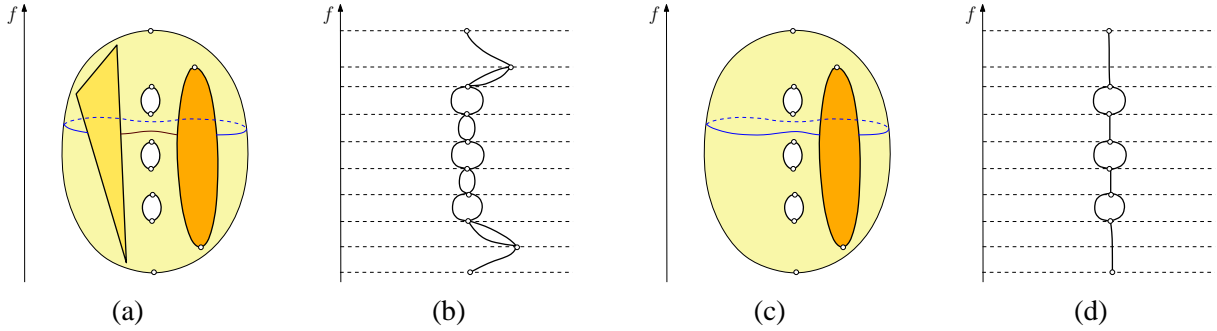


Figure 4: (a) shows a genus- g torus with the two caps missing; $g = 3$ in this case. Darker color regions indicate the two holes (missing caps) on this torus. Its Reeb graph w.r.t. the height function is shown in (b). Now if we fill the left triangle, as shown in (c), then $\Theta(g)$ number of independent vertical homological classes become horizontal, thus killing $\Theta(g)$ number of loops in the Reeb graph, which is shown in (d). In other words, by adding just one simplex (a triangle), the first betti number decreases by $\Theta(g)$.

Consistent base cycles. From now on, let $\mathbf{G}^{(i)}$ denote the cycle-basis of $H_1(R_i)$ that we maintain at the i -th step. For each cycle $\gamma \in \mathbf{G}^{(i)}$, we associate with it a *birth-time* $t(\gamma)$, which is the earliest time (index) $k \leq i$ such that some pre-image of the homology class $[\gamma]$ under the map $\xi_*^{k,i} : H_1(R_k) \rightarrow H_1(R_i)$ exists. In order to extract $\beta^{i,j}$, we wish to maintain the following *consistency condition* between $\mathbf{G}^{(i)}$ and $\mathbf{G}^{(j)}$: let $\mathbf{G}^{(i)} = \{\alpha_1^{(i)}, \alpha_2^{(i)}, \dots, \alpha_r^{(i)}\}$ and $\mathbf{G}^{(j)} = \{\alpha_1^{(j)}, \dots, \alpha_s^{(j)}\}$. Consider the set \widehat{G} of images of cycles $\{\alpha_l^{(i)}\}$ in R_j . $\mathbf{G}^{(i)}$ and $\mathbf{G}^{(j)}$ are *consistent* if the cardinality of $\widehat{G} \cap \mathbf{G}^{(j)}$ is *exactly* $\beta^{i,j}$. Notice that there are always $\beta^{i,j}$ number of independent cycles in \widehat{G} . However, its intersection with $\mathbf{G}^{(j)}$ may have much smaller cardinality. A sequence of cycle-bases $\{\mathbf{G}^{(i)} \mid i \in [1, n]\}$ is *consistent* if the consistency condition holds for any pair $\mathbf{G}^{(i)}$ and $\mathbf{G}^{(j)}$, $0 \leq i < j \leq n$. The following claim implies that we can read off $\beta^{i,j}$ easily from a consistent sequence of cycle-bases.

Claim 5.1 *If a sequence of cycle-bases $\{\mathbf{G}^{(i)} \mid i \in [1, n]\}$ is consistent, then for any $1 \leq i < j \leq n$, $\beta^{i,j}$ equals the number of cycles in $\mathbf{G}^{(j)}$ whose birth-time is smaller than or equal to i .*

Proof: Consider a pair of indices $i < j$ and the corresponding cycle-basis $\mathbf{G}^{(i)}$ for $H_1(R_i)$ and $\mathbf{G}^{(j)}$ for $H_1(R_j)$. Assume that there are k cycles in $\mathbf{G}^{(j)}$ with birth-time smaller than or equal to i . Since all these cycles are independent in R_j (and thus in R_i), we have that $k \leq \beta^{i,j}$. On the other hand, since $\mathbf{G}^{(i)}$ and $\mathbf{G}^{(j)}$ are consistent, we have that $k \geq \beta^{i,j}$, implying that $k = \beta^{i,j}$. ■

Algorithm description. In light of Claim 5.1, our goal is to maintain consistent cycle-bases at any moment. We now describe how we update the set of base cycles as we move from K_k to $K_{k+1} = K_k \cup \{\sigma\}$; σ can be a 0-, 1-, or 2-simplex. Set $g_i := \text{rank}(H_1(R_i))$ for any $i \in [1, n]$. Assume at k -th step we already have consistent $\{\mathbf{G}^{(i)} \mid i \in [1, k]\}$. For each cycle-basis $\mathbf{G}^{(i)}$, we also maintain the birth-time of each cycle in it. Assume cycles in $\mathbf{G}^{(k)} = \{\gamma_1, \dots, \gamma_{g_k}\}$ are sorted by their birth-times. At the beginning of the k -th step, we first use the incremental algorithm from [21] to compute the Reeb graph R_{k+1} from R_k . We next need to update $\mathbf{G}^{(k)}$ to $\mathbf{G}^{(k+1)}$ for R_{k+1} so that $\mathbf{G}^{(k+1)}$ is consistent with each $\mathbf{G}^{(i)}$ for $i \in [1, k]$. There are three cases.

Case 1: σ is a vertex. A new connected component is created in K_{k+1} , consisting of only σ . Similarly, a new node is created in R_{k+1} . The set of base 1-cycles are not affected, and $\mathbf{G}^{(k+1)} = \mathbf{G}^{(k)}$.

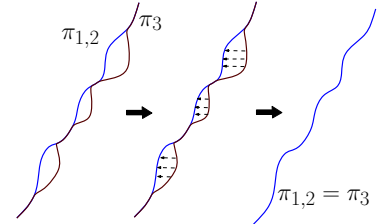
Case 2: $\sigma = pq$ is an edge. Let $\hat{p} = \Phi_k(p)$ and $\hat{q} = \Phi_k(q)$ be the images of endpoints p and q of σ in the Reeb graph R_k . Adding σ to K_k creates a new edge $e = \hat{p}\hat{q}$ in R_{k+1} . If \hat{p} and \hat{q} are not in the same connected component in R_k , then adding e will only reduce the rank of $H_0(R_k)$ by 1 and does not affect $H_1(R_k)$. In that case $\mathbf{G}^{(k+1)} = \mathbf{G}^{(k)}$. Otherwise, \hat{p} and \hat{q} are already connected in R_k . Adding e results in $\text{rank}(H_1(R_{k+1})) = \text{rank}(H_1(R_k)) + 1$. Let γ be any cycle in R_{k+1} that contains e (which can be computed easily in linear time). All previous base cycles in $\mathbf{G}^{(k)}$ will remain independent in R_{k+1} , and we simply set $\mathbf{G}^{(k+1)} = \mathbf{G}^{(k)} \cup \{\gamma\}$. The birth-time for γ is $k + 1$.

Case 3: σ is a triangle. The first two cases are simple and similar to the cases of standard persistence algorithm. Case 3 is much more complicated. In particular, unlike the standard persistence algorithm wherein adding a triangle may reduce β_1 by at most 1, the rank of $H_1(R_k)$ may decrease by $\Theta(g_k)$. What happens is that even though $\beta_1(K_k)$ reduces by at most 1, arbitrary number of vertical homology classes can be converted into horizontal homology classes. An example is given in Figure 4.

Let $\sigma = \Delta pqr$, and let $\hat{p} = \Phi_k(p)$, $\hat{q} = \Phi_k(q)$ and $\hat{r} = \Phi_k(r)$ be the images of the three endpoints of σ in R_k , respectively. Assume without loss of generality that $f(p) \leq f(q) \leq f(r)$ and set $e_1 = pq$, $e_2 = qr$ and $e_3 = pr$. First, we compute the image of each e_i in R_k , which is necessarily a monotone path (i.e, monotonic in function values) denoted by $\pi_i = \Phi_k(e_i)$. These images can be computed in $O(n_v)$ time using the incremental algorithm and the data structure of [21]. By our assumption of $f(p) \leq f(q) \leq f(r)$, π_1 and π_2 are disjoint in their interiors, while π_3 may share subcurves with π_1 and π_2 . Set $\pi_{1,2} := \pi_1 \circ \pi_2$ to be the concatenation of π_1 and π_2 , which is still a monotone path, and note $\pi_{1,2}$ and π_3 share the same two endpoints.

Now if $\pi_{1,2}$ and π_3 coincide in R_k , the addition of triangle σ does not ensue any change, that is, $R_{k+1} = R_k$ and $\mathbf{G}^{(k+1)} = \mathbf{G}^{(k)}$. In this case, the vertical homology of K_k remains the same; either σ destroys a horizontal homology class in $H_1(K_k)$, or it creates a 2-cycle.

Otherwise, the H_1 -homology of the Reeb graph changes. Assume the two monotone paths $\pi_{1,2}$ and π_3 form s simple loops between them (see the right figure where $s = 3$). Then, with the addition of σ , each point in π_3 is mapped to the corresponding point in $\pi_{1,2}$ with the same function value. Hence this process collapses all these s independent loops and we have $g_{k+1} = g_k - s$.



We now describe how to compute $\mathbf{G}^{(k+1)}$ for this case. First, we need to compute the image $\hat{G} := \xi_k(\mathbf{G}^{(k)})$ of the set of base cycles $\mathbf{G}^{(k)}$ in R_{k+1} . To do this, we need the map ξ_k . Observe that ξ_k maps each edge in R_k either to the same edge in R_{k+1} , or to a monotone path in R_{k+1} . The latter case can potentially happen only for edges in the paths $\pi_{1,2}$ and π_3 — in particular, for those edges in subcurves from $\pi_{1,2}$ and π_3 that are merged together. Since both $\pi_{1,2}$ and π_3 are monotone, images of edges from $\pi_{1,2}$ and π_3 can be computed in $O(|\pi_{1,2}| + |\pi_3|) = O(n_v)$ time by merging the sorted lists of vertices in $\pi_{1,2}$ and π_3 . Hence we can compute the map ξ_k in $O(n_v)$ time.

Once ξ_k is computed, given a simple cycle γ from R_k , we can compute its image in R_{k+1} in $O(n_v)$ time. This is because (i) there are $O(n_v)$ number of edges in γ ; and (ii) the total size of the images of edges from γ in R_{k+1} has an upper bound $|\gamma| + |\xi_k(\pi_{1,2})| + |\xi_k(\pi_3)| = O(n_v)$. The set of cycles $\hat{G} := \xi_k(\mathbf{G}^{(k)})$ in R_{k+1} can then be computed in $O(n_v g_k)$ time. Let $\hat{G} = \{\hat{\gamma}_1, \dots, \hat{\gamma}_{g_k}\}$.

The remaining task is to construct $\mathbf{G}^{(k+1)}$ that is consistent with $\mathbf{G}^{(i)}$ for any $i \leq k$. One needs $g_{k+1} = \text{rank}(\xi_*^{k,k+1})$ independent cycles from \hat{G} to make $\mathbf{G}^{(k+1)}$ consistent with $\mathbf{G}^{(k)}$. To this end, we perform the following two steps.

- (S1). We represent each cycle in \widehat{G} as a linear combination of cycles in a basis for the graph R_{k+1} .
- (S2). We check the dependency of cycles in \widehat{G} in order of their birth-times, and remove redundant cycles to obtain $\mathbf{G}^{(k+1)}$.

Step (S1). Since R_{k+1} is a graph, we compute a canonical basis of cycles, $B = \{\alpha_1, \dots, \alpha_{g_{k+1}}\}$, in the following standard way. Construct an arbitrary spanning tree T of R_{k+1} . Let $E = \{e_1, \dots, e_{g_{k+1}}\}$ denote the set of non-tree edges in R_{k+1} . Each edge $e_i = pq \in E$ creates a canonical cycle that concatenates edge e_i with the two unique paths in T from p and q to their common ancestor. We set α_i to be this canonical cycle created by e_i . Obviously, each e_i appears exactly once among all cycles in B . Given a cycle $\gamma \in \widehat{G}$, we need to find coefficients c_i s such that $\gamma = \sum_{i=1}^{g_{k+1}} c_i \alpha_i$, where each c_i is either 0 or 1. Since e_i appears only in α_i , we have c_i equal the number of times e_i appears in γ modulo 2. Since γ is a simple curve, c_i is 1 if $e_i \in \gamma$ and 0 otherwise. Hence all c_i s for $i \in [1, g_{k+1}]$ can be computed in $O(n_v)$ time for one curve γ . Computing the coefficients of all cycles in \widehat{G} takes $O(n_v g_k)$ time.

Step (S2). Recall that cycles in $\mathbf{G}^{(k)} = \{\gamma_1, \dots, \gamma_{g_k}\}$ are sorted by increasing order of their birth-times. Note, the birth-time of the cycle $\widehat{\gamma}_i \in \widehat{G}$, which is the image of the cycle $\gamma_i \in \mathbf{G}^{(k)}$ in R_{k+1} , may be smaller than the birth-time of γ_i . Now represent cycles in \widehat{G} with respect to the canonical basis $B = \{\alpha_1, \dots, \alpha_{g_{k+1}}\}$ in a matrix M , where the i th column of M , denoted by $\text{col}_M[i]$, contains the coordinates of $\widehat{\gamma}_i$ under basis B ; that is, $\widehat{\gamma}_i = \sum_{j=1}^{g_{k+1}} \text{col}_M[i][j] \alpha_j$. Obviously, the matrix M has size $g_k \times g_{k+1}$.

Next, we perform a left-to-right reduction of matrix M , which is the same as the reduction of the adjacency matrix used in the standard persistence algorithm [7, 16]. In particular, the only operation that one can use is to add a column to another one on its right. For a column $\text{col}_M[i]$, let its *low-row index* denote the largest index j such that $\text{col}_M[i][j] = 1$. At the end of the reduction, each column is either empty or has a unique low-row index; that is, no other column can have the same low-row index as this one. We set $\mathbf{G}^{(k+1)}$ as the subset of \widehat{G} whose corresponding columns in the reduced matrix M' is not all zeros. The reduction takes time $O(g_{k+1} g_k^2)$. Intuitively, the consistency of $\mathbf{G}^{(k+1)}$ with each $\mathbf{G}^{(i)}$ for $i \in [1, k]$ follows from the left-to-right reduction. It guarantees that if a set of cycles in \widehat{G} are dependent, then only those created earlier (i.e, with smaller birth-time) will be kept.

Lemma 5.2 $\mathbf{G}^{(k+1)}$ as constructed above provides a basis of $H_1(R_{k+1})$. Furthermore, if $\{\mathbf{G}^{(1)}, \dots, \mathbf{G}^{(k)}\}$ is consistent, so is $\{\mathbf{G}^{(1)}, \dots, \mathbf{G}^{(k+1)}\}$.

Proof: Let M' denote the reduced matrix of M . Recall that $\widehat{G} = \{\widehat{\gamma}_1, \dots, \widehat{\gamma}_{g_k}\}$ contains the images of cycles from $\mathbf{G}^{(k)}$ in R_{k+1} . Set $\widehat{G}_i = \{\widehat{\gamma}_1, \dots, \widehat{\gamma}_i\}$ and let G'_i be the set of cycles from \widehat{G}_i whose corresponding column in the reduced matrix M' is non-empty (i.e, not all zeros). In other words, $G'_i = \widehat{G}_i \cap \mathbf{G}^{(k+1)}$ is the intersection between \widehat{G}_i and the set $\mathbf{G}^{(k+1)}$ constructed by our algorithm. By induction on i , it is easy to show that for any $i \in [1, g_k]$, cycles in G'_i generate the same subgroup of $H_1(R_{k+1})$ as \widehat{G}_i . It then follows that, in the end, cycles in $\mathbf{G}^{(k+1)} = G'_{g_k}$ are all independent in R_{k+1} and $|\mathbf{G}^{(k+1)}|$ equals the rank of the homology group generated by cycles in \widehat{G} , which is $\beta^{k,k+1} = g_{k+1}$. This proves the first part of the claim.

For the second part of the claim, first note that $\mathbf{G}^{(k+1)}$ is consistent with $\mathbf{G}^{(k)}$ as $\widehat{G} \cap \mathbf{G}^{(k+1)} = \mathbf{G}^{(k+1)}$ and has cardinality g_{k+1} . Now consider an arbitrary $\mathbf{G}^{(i)}$ with $i < k$. Since $\{\mathbf{G}^{(1)}, \dots, \mathbf{G}^{(k)}\}$ are consistent, and cycles $\{\gamma_1, \dots, \gamma_{g_k}\}$ in $\mathbf{G}^{(k)}$ are sorted by their birth-times, it follows from Claim 5.1 that the first $s = \beta^{i,k}$ number of cycles $G_s = \{\gamma_1, \dots, \gamma_s\}$ from $\mathbf{G}^{(k)}$ are images of cycles from $\mathbf{G}^{(i)}$. Hence the image of cycles from $\mathbf{G}^{(i)}$ in R_{k+1} are exactly the cycles in \widehat{G}_s , and classes of cycles in \widehat{G}_s generate the persistent homology group $\xi_*^{i,k+1}(H_1(R_f(K_i)))$. On the other hand, as mentioned above, classes of cycles in $G'_s = \widehat{G}_s \cap \mathbf{G}^{(k+1)}$ generate the same subgroup of $H_1(R_{k+1})$ as \widehat{G}_s . Since cycles in G'_s are independent, G'_s has rank $\beta^{i,k+1}$, implying that $\mathbf{G}^{(k+1)}$ is consistent with $\mathbf{G}^{(i)}$, for any $i \in [1, k]$. The second part of the claim then follows. \blacksquare

Finally, for our algorithm to continue into the next iteration, we also need to maintain the birth-times for each cycle in $\mathbf{G}^{(k+1)}$. This is achieved by the following claim.

Claim 5.3 *Let $\mathbf{G}^{(k+1)} = \{\widehat{\gamma}_{I_1}, \dots, \widehat{\gamma}_{I_{g_{k+1}}}\}$, where I_i s are the set of indices of non-zero columns in the reduced matrix M' . Then the birth-time of $\widehat{\gamma}_{I_i}$ equals to the birth-time of γ_{I_i} for any $i \in [1, g_{k+1}]$.*

Proof: Recall that $\mathbf{G}^{(k+1)}$ contains the set of cycles $\widehat{\gamma}_{I_i}$ where $\{I_i\}$ is the set of indices of non-zero columns from the reduced matrix M' . Given a cycle $\alpha \in \mathbf{G}^{(i)}$, let $\text{birthtime}(\alpha)$ denote the birth-time of α . Assume that one of the cycles, say $\widehat{\gamma}_m \in \mathbf{G}^{(k+1)}$, has a birth-time that is different from that of $\gamma_m \in \mathbf{G}^{(k)}$. Set $t := \text{birthtime}(\widehat{\gamma}_m)$. Since $\widehat{\gamma}_m = \xi^k(\gamma_m)$, we have $t \leq \text{birthtime}(\gamma_m)$. Since the two birth-times are different, t must be strictly smaller than the birth-time of γ_m .

Furthermore, there exists a cycle $\alpha \in R_t$ such that its image $\alpha_1 := \xi^{t,k}(\alpha)$ in R_k is not homologous to γ_m , while its image $\alpha_2 := \xi^{t,k+1}(\alpha)$ in R_{k+1} is $\widehat{\gamma}_m$. On the other hand, α_1 can be uniquely written as a linear combination of a subset of cycles from $\mathbf{G}^{(k)}$, say $\alpha_1 = \gamma_{J_1} + \dots + \gamma_{J_t}$. It is easy to verify that the birth-time of each γ_{J_i} is at most t . Since $t < \text{birthtime}(\gamma_m)$, it follows that all indices J_i s are strictly smaller than m (as cycles in $\mathbf{G}^{(k)}$ are sorted by their birth-times). However, this is not possible since the resulting m -th column will be all zero at the time when we reduce the m -th column to construct $\mathbf{G}^{(k+1)}$ as $\widehat{\gamma}_m = \sum_i \widehat{\gamma}_{J_i}$. Hence the cycle $\widehat{\gamma}_m$ cannot be chosen as a base cycle in $\mathbf{G}^{(k+1)}$ reaching a contradiction. It follows that $t = \text{birthtime}(\gamma_m)$, or more generally, $\text{birthtime}(\widehat{\gamma}_{I_i}) = \text{birthtime}(\gamma_{I_i})$ for every index I_i of non-zero column in the reduced matrix M' . ■

Putting everything together, we conclude with the following main result.

Theorem 5.4 *Given a filtration $K_1 \subset \dots \subset K_n$ of a simplicial complex K_n with a piecewise linear function $f : K_n \rightarrow \mathbb{R}$, we can compute all persistent first betti numbers for the induced sequence of Reeb graphs $R_f(K_i)$ s in $O(\sum_{i=1}^n (n_v g_i + g_i^3)) = O(nn_e^3)$ time, where n_v and n_e are the number of vertices and edges in K_n , respectively, n is the size of 2-skeleton of K_n , and g_i is the first betti number of the Reeb graph $R_f(K_i)$.*

6 Conclusions and discussions

In this paper, we present a simple and efficient algorithm to approximate the Reeb graph $R_f(M)$ of a map $f : M \rightarrow \mathbb{R}$ from point data sampled from a smooth and compact manifold M . Given that Reeb graph is an abstract graph with a function defined on it, we only approximate its topology together with the range information for each loop in it. It will be interesting to see whether the Reeb graph we compute from the point data is also geometrically close to some specific embedding of the Reeb graph $R_f(M)$ in the hidden domain M . To this end, our results in Section 4.1 on mappings between cycles can be useful.

We also study how to compute the ‘‘persistence’’ of loops in a Reeb graph by measuring their life time as the defining domain grows. An immediate question is to see whether the time complexity can be further improved to match that of the standard persistence algorithm in the worst case.

Finally, it will be interesting to explore whether one can leverage the simple structure and efficient computation of the Reeb graph to retrieve topological information for various spaces efficiently. For example, given a 3-manifold with a function f defined on it, its vertical H_1 -homology is already encoded in the Reeb graph and can thus be computed in near-linear time. Can we retrieve the horizontal H_1 -homology efficiently by tracking the levelsets of f , or by defining another function that is somewhat ‘‘orthogonal’’ to f ?

Acknowledgments.

The authors thank anonymous reviewers for helpful comments. This work is supported by National Science Foundation under grants CCF-0915996, CCF-1116258 and CCF-1048983.

References

- [1] M. Bernstein, V. de Silva, J. Lanford, and J. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Dept. Psychology, Stanford University, USA, 2000.
- [2] S. Biasotti, D. Giorgi, M. Spagnuolo, and B. Falcidieno. Reeb graphs for shape analysis and applications. *Theor. Comput. Sci.*, 392(1-3):5–22, 2008.
- [3] F. Chazal, L. J. Guibas, S. Oudot, and P. Skraba. Analysis of scalar fields over point cloud data. In *Proc. 20th ACM-SIAM Sympos. Discrete Algorithms*, pages 1021–1030, 2009.
- [4] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba. Persistence-based clustering in Riemannian manifolds. In *Proc. 27th Annu. Sympos. Comput. Geom.*, pages 97–106, 2011.
- [5] F. Chazal and S. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Sympos. on Comput. Geom.*, pages 232–241, 2008.
- [6] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Extending persistence using Poincaré and Lefschetz duality. *Foundations of Computational Mathematics*, 9(1):79–103, 2009.
- [7] D. Cohen-Steiner, H. Edelsbrunner, and D. Morozov. Vines and vineyards by updating persistence in linear time. In *Proc. 22nd Annu. Sympos. Comput. Geom.*, pages 119–126, 2006.
- [8] K. Cole-McLaughlin, H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci. Loops in Reeb graphs of 2-manifolds. *Discrete Comput. Geom.*, 32(2):231–244, 2004.
- [9] T. K. Dey. *Curve and Surface Reconstruction: Algorithms with Mathematical Analysis*. Cambridge U. Press, New York, NY, USA, 2007.
- [10] T. K. Dey and K. Li. Cut locus and topology from surface point data. In *Proc. 25th Annu. Sympos. Comput. Geom.*, pages 125–134, 2009.
- [11] T. K. Dey, J. Sun, and Y. Wang. Approximating loops in a shortest homology basis from point data. In *Proc. 26th Annu. Sympos. Compu. Geom.*, pages 166–175, 2010.
- [12] H. Doraiswamy and V. Natarajan. Efficient output-sensitive construction of Reeb graphs. In *Proc. 19th Internat. Sym. Alg. and Comput.*, pages 556–567, 2008.
- [13] H. Doraiswamy and V. Natarajan. Efficient algorithms for computing Reeb graphs. *Computational Geometry: Theory and Applications*, 42:606–616, 2009.
- [14] H. Edelsbrunner and J. Harer. Persistent homology — a survey. In J. E. Goodman, J. Pach, and R. Pollack, editors, *Surveys on Discrete and Computational Geometry. Twenty Years Later*, pages 257–282. Amer. Math. Soc., Providence, Rhode Island, 2008. Contemporary Mathematics 453.
- [15] H. Edelsbrunner, J. Harer, A. Mascarenhas, V. Pascucci, and J. Snoeyink. Time-varying Reeb graphs for continuous space-time data. *Comput. Geom.*, 41(3):149–166, 2008.

- [16] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.*, 28:511–533, 2002.
- [17] W. Harvey, R. Wenger, and Y. Wang. A randomized $O(m \log m)$ time algorithm for computing Reeb graph of arbitrary simplicial complexes. In *Proc. 26th Annu. Sympos. Compu. Geom.*, pages 267–276, 2010.
- [18] J.-C. Hausmann. On the Vietoris-Rips complexes and a cohomology theory for metric spaces. In *Prospects in Topology: Proc. Conf. in Honour of William Browder*, Ann. Math. Stud., 138, pages 175–188. Princeton Univ. Press, 1995.
- [19] J. R. Munkres. *Elements of Algebraic Topology*. Westview Press, 1996.
- [20] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39:419–441, 2008.
- [21] V. Pascucci, G. Scorzelli, P.-T. Bremer, and A. Mascarenhas. Robust on-line computation of Reeb graphs: simplicity and speed. *ACM Trans. Graph.*, 26(3):58, 2007.
- [22] Y. Shinagawa and T. L. Kunii. Constructing a Reeb graph automatically from cross sections. *IEEE Comput. Graph. Appl.*, 11(6):44–51, 1991.
- [23] J. Tierny, A. Gyulassy, E. Simon, and V. Pascucci. Loop surgery for volumetric meshes: Reeb graphs reduced to contour trees. *IEEE Trans. Vis. Comput. Graph.*, 15(6):1177–1184, 2009.

A Proof for Lemma 4.2

The second part of claim (ii) follows immediately from Claim 4.1. The first part of claim (ii) follows from claim (i). Hence we focus on proving claim (i). First, we show the following two claims: (C-1) given a cycle α from $\mathbf{Z}_1(M)$, $[\alpha] = [h_{\#}(d(\alpha))]$; and (C-2) given a cycle $\hat{\alpha} \in \mathbf{Z}_1(\mathcal{R}^r(P))$, we have that $[\hat{\alpha}] = [d(h_{\#}(\hat{\alpha}))]$. Note that cycles in $\mathbf{Z}_1(\mathcal{R}^r(P))$ are simplicial cycles, while cycles in $\mathbf{Z}_1(M)$ are singular cycles.

Proving (C-1). We show that given a cycle α from $\mathbf{Z}_1(M)$, $[\alpha] = [h_{\#}(d(\alpha))]$. Let $\gamma \subseteq M$ be the carrier of α . To map α to $d(\alpha)$, suppose that its carrier γ is broken into k pieces as described earlier using the Decomposition method. For the i -th piece with endpoints x_i and x_{i+1} , let p_i and p_{i+1} be their closest point in P , respectively; recall that $d(\alpha)$ is the concatenation of all edges $p_i p_{i+1}$ for $i \in [1, k]$.

Now consider the cycle $h_{\#}(d(\alpha))$ in M : its carrier γ' is the concatenation of $h(p_i p_{i+1}) \subset M$ for all $i \in [1, k]$. By Proposition 3.3 of [11], each curve $h(p_i p_{i+1})$ has endpoints p_i and p_{i+1} , and it is contained in the union of the two Euclidean balls of radius r centered at p_i and at p_{i+1} . Since p_i and p_{i+1} are within r Euclidean distance, $h(p_i p_{i+1})$ is contained in the Euclidean balls of radius $2r$ centered at p_i and at p_{i+1} . Notice that the geodesic distance and the Euclidean distance between two points $x, y \in M$ approximate each other when x and y are close enough (see e.g. Proposition 1.2 from [11]). It follows that when r is smaller than $\rho(M)/4$, $h(p_i p_{i+1})$ is contained in both geodesic balls of radius $3r$ centered at $p_i \in M$ and $p_{i+1} \in M$.

Let $\gamma[x_i, x_{i+1}]$ denote the subcurve of γ from x_i to x_{i+1} . Since the length of $\gamma[x_i, x_{i+1}]$ is less than r by construction, the curve $\gamma[x_i, x_{i+1}]$ is contained in the geodesic balls of radius r centered at x_i and x_{i+1} . This implies that the curve $\gamma[x_i, x_{i+1}]$ is contained in the *geodesic tubular neighborhood*

$$\text{Tub}_r(\pi_g(x_i, x_{i+1})) := \{y \in M \mid d(y, \pi_g(x_i, x_{i+1})) \leq r\}$$

where $\pi_g(x, y)$ denote a minimizing geodesic between two points $x, y \in M$. By Proposition 3.7 of [11], $\text{Tub}_r(\pi_g(x_i, x_{i+1}))$ is contractible and hence $\gamma[x_i, x_{i+1}]$ is homotopy equivalent to $\pi_g(x_i, x_{i+1})$.

On the other hand, due to the sampling condition, the geodesic distances between x_i and p_i , and between x_{i+1} and p_{i+1} , are both bounded by r . Combining this with the fact that $h(p_i p_{i+1})$ lies within the geodesic balls of radius $3r$ centered at both p_i and p_{i+1} , we have that any point in $h(p_i p_{i+1})$ is within geodesic distance $4r$ to both x_i and to x_{i+1} . Hence $h(p_i p_{i+1})$ lies within the geodesic tubular neighborhood $\text{Tub}_{4r}(\pi_g(x_i, x_{i+1}))$. Again by Proposition 3.7 of [11], when $r \leq \rho_c(M)/4$, the curve $\pi_g(x_i, p_i) \circ h(p_i p_{i+1}) \circ \pi_g(p_{i+1}, x_{i+1})$ is homotopy equivalent to $\pi_g(x_i, x_{i+1})$ and thus homotopy equivalent to $\gamma[x_i, x_{i+1}]$. In fact, one can find a homotopy h_i that keeps p_i and p_{i+1} on the geodesics $\pi_g(x_i, p_i)$ and $\pi_g(x_{i+1}, p_{i+1})$ respectively so that two maps h_i and h_{i+1} , for $i \in [1, k-1]$, are consistent in mapping the common endpoints x_{i+1} . Therefore, we can combine all such h_i 's to obtain a homotopy between γ' (which is the carrier of $h_{\#}(d(\alpha))$) and γ (which is the carrier of α). It follows that $[h_{\#} \circ d(\alpha)] = [\alpha]$.

Proving (C-2). We now show that given any (simplicial) cycle $\hat{\alpha} \in \mathbf{Z}_1(\mathcal{R}^r(P))$, we have that $[\hat{\alpha}] = [d \circ h_{\#}(\hat{\alpha})]$. First, consider the image $\gamma = h(\hat{\alpha})$ of $\hat{\alpha}$ in M ; γ is the carrier of the cycle $h_{\#}(\hat{\alpha}) \in \mathbf{Z}_1(M)$. By construction, γ is the concatenation of $h(e)$ s for every edge $e = pq$ in the simplicial cycle $\hat{\alpha}$. By Proposition 3.3 in [11], each curve $h(e)$ is contained inside $M \cap (B_r(p) \cup B_r(q))$. Hence, for any point $x \in h(e)$, its geodesic distance to p and to q is bounded by $3r$.

Now consider mapping the cycle $h_{\#}(\hat{\alpha})$ carried by γ back to $\mathcal{R}^r(P)$ using the decomposition method described earlier. Consider the set of breaking points x_i 's in the subcurve $h(e) \subset \gamma$ — assume for simplicity that the endpoints of $h(e)$, that is p and q , are also break points. Each break point x_i in $h(e)$ will be mapped to its nearest point $p_i \in P$ and the geodesic distance between x_i and p_i is at most r . Hence p_i is within $3r + r = 4r$ geodesic distance to both endpoints p and q of the edge $e \subset \hat{\alpha}$. This means that both $p_i p$ and $p_i q$ are edges in the Rips complex $\mathcal{R}^{4r}(P)$. Hence the concatenation of arcs $p_i p_{i+1}$ is homotopy equivalent to the edge e in the simplicial complex $\mathcal{R}^{4r}(P)$. Combining this homotopy equivalent map for every edge $e \in \hat{\alpha}$, we have that $d(h_{\#}(\hat{\alpha}))$ is homotopy equivalent, and thus homologous, to $\hat{\alpha}$ in $\mathcal{R}^{4r}(P)$. Finally, since the inclusion map from $\mathcal{R}^r(P)$ to $\mathcal{R}^{4r}(P)$ induces an isomorphism in the first homology groups when r is small, we have that $d(h_{\#}(\hat{\alpha}))$ is homologous to $\hat{\alpha}$ in $\mathcal{R}^r(P)$ as well. Thus $[\hat{\alpha}] = [d(h_{\#}(\hat{\alpha}))]$.

Proving Claim (i). (C-1) and (C-2) imply that d is the homology-inverse of the map $h_{\#}$. The map $h_{\#}$ induces an isomorphism $h_* : H_1(\mathcal{R}^r(P)) \rightarrow H_1(M)$. Hence d also induces an isomorphism in the homology groups $H_1(M) \rightarrow H_1(\mathcal{R}^r(P))$.

Proving Claim (iii). Claim (iii) follows easily from the constructions of d and h . In particular, consider a cycle $\alpha \in \mathbf{Z}_1(M)$ and $d(\alpha) \in \mathbf{Z}_1(\mathcal{R}^r(P))$. (The case for $\hat{\alpha}$ from $\mathbf{Z}_1(\mathcal{R}^r(P))$ and $h_{\#}(\hat{\alpha})$ from $\mathbf{Z}_1(M)$ can be similarly argued.) Let $\gamma \subset M$ be the carrier of α . The Decomposition method breaks γ into k pieces $\gamma(x_i, x_{i+1})$ s, for $i \in [0, k]$. Each piece $\gamma(x_i, x_{i+1})$ is mapped to the edge $p_i p_{i+1}$ where p_i is the closest point of x_i in P . Since P is an ε -sample of M , and since the length of $\gamma(x_i, x_{i+1})$ is at most $r - 2\varepsilon$, any point x in $\gamma(x_i, x_{i+1})$ is within r geodesic distance to the point $p_i \in M$. Hence by the Lipschitz condition of f , we have $|f(x_i) - f(p_i)| \leq r \cdot \text{Lip}_f$. It follows that

$$\begin{aligned} [\min_{i \in [0, k]} f(p_i) + r \cdot \text{Lip}_f, \max_{i \in [0, k]} f(p_i) - r \cdot \text{Lip}_f] &\subseteq \text{range}(\gamma) = \text{range}(\alpha) \\ &\subseteq [\min_{i \in [0, k]} f(p_i) - r \cdot \text{Lip}_f, \max_{i \in [0, k]} f(p_i) + r \cdot \text{Lip}_f]. \end{aligned}$$

On the other hand, note that under \mathbb{Z}_2 coefficient, $\text{range}(d(\alpha)) \subseteq [\min_{i \in [0, k]} f(p_i), \max_{i \in [0, k]} f(p_i)]$ (and it can be much smaller than this interval). It then follows that $\text{range}(d(\alpha))$ is onside- $(r \cdot \text{Lip}_f)$ -close to $\text{range}(\alpha)$.

Proof of Claim (iv). Consider any homology class $\omega \in H_1(M)$. By Claim (iii) we have that the range of $d_*(\omega)$ is onside- $(r \cdot \text{Lip}_f)$ -close to the range of ω : indeed, choose the thinnest cycle α of ω , we have $\text{range}(d(\alpha))$ is onside- $(r \cdot \text{Lip}_f)$ -close to $\text{range}(\alpha) = \text{range}(\omega)$. Since $\text{range}(d_*(\alpha)) \subseteq \text{range}(d(\gamma))$, $\text{range}(d_*(\alpha))$ is also onside- $(r \cdot \text{Lip}_f)$ -close to $\text{range}(\omega)$.

Now map $d_*(\omega)$ back to $H_1(M)$, we have that the range of $h_*(d_*(\omega))$ is also onside- $(r \cdot \text{Lip}_f)$ -close to the range of $d_*(\omega)$ by Claim (iii). Since $h_*(d_*(\omega)) = \omega$, the ranges of ω and $d_*(\omega)$ are $(r \cdot \text{Lip}_f)$ -Hausdorff-close.

The statement for $\hat{\omega}$ and $h_*(\hat{\omega})$ can be argued similarly.